



Universidad  
Carlos III de Madrid

Departamento de Informática

## PROYECTO FIN DE CARRERA

INGENIERÍA TÉCNICA EN INFORMÁTICA DE GESTIÓN

# Auditoría de calidad de bases de datos de carácter personal

Autor: Alejandro Viana Lara

Tutor: Miguel Ángel Ramos

Leganés, curso 2012-2013



PAGINA DEJADA EN BLANCO INTENCIONADAMENTE



Título: Auditoría de calidad de bases de datos de carácter personal

Autor: Alejandro Viana Lara

Tutor: Miguel Ángel Ramos

Director / Supervisor: Antonio García Carmona

## EL TRIBUNAL

Presidente: Ana Isabel González-Tablas

Vocal: Jorge Blasco

Secretario: Óscar Pérez

Realizado el acto de defensa y lectura del Proyecto Fin de Carrera el día 28 de febrero de 2013 en Leganés, en la Escuela Politécnica Superior de la Universidad Carlos III de Madrid, acuerda otorgarle la CALIFICACIÓN de NOTABLE (8).

SECRETARIO

VOCAL

PRESIDENTE



# Agradecimientos

Quisiera aprovechar la oportunidad que se me brinda para agradecer a todas esas personas que me han apoyado tanto durante este tiempo.

En primer lugar quiero dar mi más sincero agradecimiento a Antonio García, por haberme dirigido este proyecto de fin de carrera. Por todas las atenciones, por el tiempo que ha perdido conmigo, y sobre todo por su apoyo. También quisiera agradecerle a Miguel Ángel Ramos su dedicación desde el momento en que me tutorizó, persona en la que le he podido observar la gran dedicación que hay que darle al mundo de la auditoría informática.

No puedo dejar de agradecer la comprensión de familiares y amigos, puesto que ellos han sido los que más me han tenido que aguantar en mis malos momentos, enfados y demás sucesos que me han acaecido a lo largo de la elaboración de este proyecto.

Y, por último, pero no por ello menos importante, a Eugenia, porque tu apoyo durante toda la carrera ha sido muy importante para mí.



# Citas célebres

*Convivir con la no calidad de datos, no gracias.*  
Anónimo.

*En la carrera por la calidad no hay línea de meta.*  
Kearns

*La calidad nunca es un accidente; siempre es el resultado de un esfuerzo de la inteligencia.*  
John Ruskin (1819-1900)

*I do not worry whether something is cheap or expensive. I only worry if it is good. If it is good enough, the public will pay you back for it.*  
(No me preocupo si algo es barato o caro. Sólo me preocupa si es bueno. Si es lo suficientemente bueno, el público pagará por ello)  
Walt Disney (1901 – 1966)

*"He ofendido a Dios y a la humanidad porque mi trabajo no tuvo la calidad que debía haber tenido".*  
Leonardo Da Vinci (1452 – 1519)

# Índice de contenidos

<b>AGRADECIMIENTOS.....</b>	<b>IV</b>
<b>CITAS CÉLEBRES .....</b>	<b>V</b>
<b>RESUMEN .....</b>	<b>2</b>
<b>ABSTRACT .....</b>	<b>3</b>
<b>BLOQUE I.....</b>	<b>4</b>
<b>1. INTRODUCCIÓN Y OBJETIVOS .....</b>	<b>4</b>
1.1 Introducción .....	4
1.2 Objetivos.....	5
1.3 Fases del desarrollo .....	6
1.4 Medios empleados.....	8
1.5 Estructura de la memoria.....	8
<b>BLOQUE II.....</b>	<b>12</b>
<b>2. ESTADO DEL ARTE .....</b>	<b>12</b>
<b>3. AUDITORÍA INFORMÁTICA.....</b>	<b>14</b>
3.1 Introducción .....	14
3.2 Auditoría informática .....	16
3.3 Auditoría interna y auditoría externa .....	17
3.4 Alcance de la auditoría informática .....	18
3.5 Síntomas de necesidad de una auditoría informática .....	19
3.6 Herramientas y técnicas para la auditoría informática .....	21
3.7 Metodología de trabajo de auditoría informática .....	23
3.8 Auditoría VS consultoría .....	23
<b>4. BASES DE DATOS .....</b>	<b>24</b>
4.1 Introducción .....	24
4.2 Bases de datos .....	24
4.3 Sistemas gestores de bases de datos .....	25
<b>5. AUDITORÍA DE LA CALIDAD.....</b>	<b>26</b>
5.1 ¿Qué es la calidad? .....	26
5.2 Objetivo de la auditoría de la calidad .....	28



## Auditoría de calidad de bases de datos de carácter personal

5.3	Auditoría interna y externa de calidad .....	29
5.4	Repercusiones de la auditoría de calidad .....	29
5.5	Ventajas de la auditoría de calidad .....	30
5.6	Beneficios de la calidad de datos.....	31
5.7	Problemas que afectan a la calidad de datos .....	32
5.8	A quién afecta la no calidad de los datos.....	33
5.9	Procesos de calidad .....	34
5.10	Factores de la calidad .....	36
5.11	Principio deontológico de la calidad .....	39
5.12	El coste de la no calidad de datos .....	41
5.13	Artículo de opinión: Las 10 causas más frecuentes de la no calidad de datos .....	42
<b>6.</b>	<b>IMPlicACIONES LEGALES .....</b>	<b>43</b>
6.1	¿Qué es la Ley Orgánica de Protección de Datos de carácter personal? .....	43
6.2	Antecedentes normativos .....	43
6.3	Ámbito de aplicación.....	44
6.4	¿Qué dice la Agencia Española de Protección de Datos sobre la calidad de datos? .....	45
6.5	Artículo 4 de la LOPD de la AEPD: calidad de datos .....	46
6.6	Diversas doctrinas y sentencias reducidas sobre la calidad de datos. ....	49
6.7	La importancia de la recogida de datos de carácter personal.....	51
6.8	Implicaciones legales a nivel europeo .....	51
<b>7.</b>	<b>SOLUCIONES COMERCIALES .....</b>	<b>52</b>
7.1	DEYDE .....	53
7.2	CAMERDATA .....	55
7.3	SCHÖBER .....	56
7.4	PITNEY BOWES.....	57
7.5	AXESOR.....	58
7.6	Soluciones comerciales en el marco internacional .....	60
7.6.1	TALEND* .....	60
7.6.2	BUSINESS DATA QUALITY .....	61
7.6.3	DATA QUALITY ONLINE .....	62
7.6.4	DATAFLUX .....	63
<b>BLOQUE III</b>	<b>.....</b>	<b>64</b>
<b>8. MÉTRICA .....</b>	<b>.....</b>	<b>64</b>
8.1	Introducción. ....	64
8.2	Grafismo utilizado. ....	67
8.3	Métrica a aplicar para las diferentes indizaciones.....	68
8.4	Penalizaciones a los resultados de la métrica. ....	75
<b>9. APLICACIÓN.....</b>	<b>.....</b>	<b>76</b>
9.1	Checklist .....	77
9.2	Servicio Web.....	84
9.2.1	Pruebas Formulario.....	85
<b>BLOQUE IV .....</b>	<b>.....</b>	<b>97</b>
<b>10. PRESUPUESTO .....</b>	<b>.....</b>	<b>97</b>
<b>11. LÍNEAS DE INVESTIGACIÓN FUTURAS .....</b>	<b>.....</b>	<b>99</b>
<b>12. CONCLUSIONES.....</b>	<b>.....</b>	<b>100</b>
<b>ANEXO .....</b>	<b>.....</b>	<b>101</b>
	Bibliografía.....	101



Glosario de términos y acrónimos .....	104
Formulario NOTA en la AEPD .....	105
<i>¿Cómo se da de alta una base de datos, o un fichero en la AEPD? Formulario</i>	
<i>NOTA .....</i>	<i>105</i>
<i>Quiénes deben notificar .....</i>	<i>106</i>
<i>Pasos a seguir para cumplimentar el fichero NOTA.....</i>	<i>107</i>
<i>Resumen estadísticas mensuales RGPD .....</i>	<i>110</i>
Resumen resoluciones LOPD.....	112
Ponderaciones de los campos de las diferentes bases de datos .....	114



# Índice de ilustraciones

Ilustración 1. Duración de cada actividad del proyecto .....	7
Ilustración 2. Calendario del desarrollo del proyecto.....	7
Ilustración 3. Bases de datos sin limpiar .....	12
Ilustración 4. Costes de la mala calidad .....	12
Ilustración 5. ISO 9001:2008 .....	27
Ilustración 6. Factores de la calidad .....	37
Ilustración 7. Evolución de COBIT (ISACA) .....	40
Ilustración 8. Principales causas de pérdida de calidad de datos .....	42
Ilustración 9. Soluciones DEYDE .....	53
Ilustración 10. Soluciones DEYDE 2. ....	54
Ilustración 11. Soluciones CAMERDATA .....	55
Ilustración 12. Soluciones SCHOBBER.....	56
Ilustración 13. Soluciones PITNEY BOWES .....	57
Ilustración 14. Soluciones AXESOR.....	58
Ilustración 15. Soluciones Internacional Talend* .....	60
Ilustración 16. Soluciones Internacional Business Data Quality .....	61
Ilustración 17. Soluciones Internacional Data Quality Online .....	62
Ilustración 18. Soluciones Internacional Dataflux.....	63
Ilustración 19. Esquema gráfico del protocolo de funcionamiento del servicio...	66
Ilustración 20. Gráfico de Kiviat.....	67
Ilustración 21. Campos de identificación de Universidad .....	69
Ilustración 22. Gráfico de identificación de Universidad .....	69
Ilustración 23. Campos de dirección de universidad .....	69
Ilustración 24. Gráfico de dirección de Universidad .....	70
Ilustración 25. Campos de identificación de Banco .....	70
Ilustración 26. Gráfico de identificación de Banco.....	71
Ilustración 27. Campos de dirección de Banco .....	71
Ilustración 28. Gráfico de dirección de Banco .....	71
Ilustración 29. Campos de identificación de Web de empleo .....	72
Ilustración 30. Gráfico de identificación de Web de empleo.....	72
Ilustración 31. Campos de dirección de Web de empleo .....	72



## Auditoría de calidad de bases de datos de carácter personal

Ilustración 32. Gráfico de dirección de Web de empleo .....	73
Ilustración 33. Campos de identificación de Videoclub .....	73
Ilustración 34. Gráfico de identificación de Videoclub .....	74
Ilustración 35. Campos de dirección de Videoclub .....	74
Ilustración 36. Gráfico de dirección de Videoclub .....	74
Ilustración 37. Gráfico del nivel de salud del formulario .....	83
Ilustración 38. Interfaz principal parte 1 .....	84
Ilustración 39. Interfaz principal parte 2 .....	84
Ilustración 40. Ejemplo formulario 1 .....	85
Ilustración 41. Ejemplo formulario 2 .....	86
Ilustración 42. Ejemplo formulario 3 .....	87
Ilustración 43. Resultado ejemplo formulario 3 - 1 .....	88
Ilustración 44. Resultado ejemplo formulario 3 - 2 .....	88
Ilustración 45. Ejemplo formulario 4 .....	89
Ilustración 46. Resultado ejemplo formulario 4 - 1 .....	90
Ilustración 47. Resultado ejemplo formulario 4 - 2 .....	90
Ilustración 48. Ejemplo formulario 5 .....	91
Ilustración 49. Resultado ejemplo formulario 5 - 1 .....	92
Ilustración 50. Resultado ejemplo formulario 5 - 2 .....	92
Ilustración 51. Ejemplo formulario 6 .....	93
Ilustración 52. Resultado ejemplo formulario 6 - 1 .....	94
Ilustración 53. Resultado ejemplo formulario 6 - 2 .....	94
Ilustración 54. Ejemplo formulario 7 .....	95
Ilustración 55. Resultado ejemplo formulario 7 - 1 .....	96
Ilustración 56. Resultado ejemplo formulario 7 - 2 .....	96
Ilustración 57. NOTA 1 .....	107
Ilustración 58. NOTA 2 .....	108
Ilustración 59. NOTA 3 .....	108
Ilustración 60. NOTA 4 .....	109
Ilustración 61. NOTA 5 .....	109
Ilustración 62. NOTA 6 .....	110
Ilustración 63. NOTA 7 .....	111
Ilustración 64. NOTA 8 .....	111
Ilustración 65. NOTA 9 .....	111



# Índice de tablas

Tabla 1. Principales procesos en los objetivos de calidad .....	34
Tabla 2. Ejemplo de checklist_1 .....	81
Tabla 3. Ejemplo de checklist_2 .....	82
Tabla 4. Nivel de salud del formulario .....	82



# Resumen

En el presente proyecto se realizará un estudio de la calidad de los datos de las bases de datos de carácter personal.

Se profundizará en la evaluación de la calidad de los datos, con respecto a su inserción, en diferentes bases de datos. Dicho tema se estudiará al ser un problema mayúsculo en la actualidad de las empresas, que genera muchas pérdidas económicas además de pérdida de información, con su correspondiente pérdida de confianza respecto al cliente final. Podremos comprobar, por ejemplo, que para formularios iguales, en diferentes bases de datos, el nivel de calidad dependerá de campos distintos.

Para ello, se propone el desarrollo de una herramienta de calidad de datos de ayuda al auditor, o a aquellas personas que tratan de mejorar o controlar la calidad, en su tarea.

Todo esto irá precedido por un análisis en profundidad de los términos de la auditoría informática, las bases de datos y de la auditoría de calidad, estudiando las implicaciones legales que la calidad conlleva, y realizando un pequeño estudio del mercado en lo referente a las empresas cuya tarea principal es el estudio de la calidad en los datos.



# Abstract

This project will study the quality of data from databases of personal data.

It will deepen in the evaluation of the quality of data with regard to its insertion, in different databases. This issue will be studied because is a major problem at present business, which generates many economic losses in addition to loss of information, with a corresponding loss of confidence in the final customer. We may see, for example, that for equal forms in different databases, the quality level will depend on different fields.

To do this, it proposes the development of a data quality tool help the auditor, or those trying to improve or control the quality in his work.

All this is preceded by a thorough analysis of the terms of the audit computer databases and the quality audit, studying the legal implications that quality implies, and doing a little market research regarding companies whose main task is the study of data quality.



# Bloque I

## Introducción y objetivos

### 1.1 Introducción

Los sistemas de bases de datos se han convertido en elementos imprescindibles en la vida cotidiana de la sociedad moderna. Cada día, la mayoría de nosotros nos encontramos con actividades que requieren algún tipo de interacción con una base de datos. Por eso, es muy importante para todas las empresas poseer un sistema de bases de datos de calidad.

Los productos de Calidad de Datos permiten definir y automatizar estándares para asegurar que los datos dentro de la organización son capturados y almacenados de tal manera que puedan ser compartidos por otras aplicaciones y sistemas en cualquier momento.

La calidad de datos es uno de los fundamentos del éxito de las organizaciones. El tener acceso a información exacta y completa es fundamental para la toma de decisiones estratégicas y de misión crítica.

Los datos con calidad apoyan y fortalecen virtualmente todas las funciones de los negocios y son especialmente importantes para las iniciativas que involucran servicios al cliente, comunicaciones y manejo de relaciones.

En cuanto al marco legal, las bases de datos que contengan datos de carácter personal, y que sean utilizadas en el plano profesional, tienen que estar amparadas bajo el marco de la L.O.P.D. respetando la legalidad vigente, y los ficheros deben de estar inscritos en la Agencia Española de Protección de datos.

Existen muchos datos desalentadores del coste que implica la falta de calidad de datos en las empresas, por ello, en este proyecto se va a hacer especial hincapié en la inserción de datos en diferentes bases de datos, y para ello se va a realizar el desarrollo de un prototipo para la ayuda del auditor en la realización de su tarea, el cual podrá ser también una herramienta o una facilidad para cualquier organización que quiera asegurar la calidad de los datos que maneja día a día.



## 1.2 Objetivos

El propósito de este proyecto es definir una métrica de valoración de la calidad de los datos introducidos en una base de datos de carácter personal.

Este proyecto nace con la pretensión de proporcionar una herramienta de ayuda al auditor, o a aquellas personas que tratan de mejorar o controlar la calidad, que permita hacer una valoración del nivel de calidad, o el nivel de salud con la que se introducen los datos en una bbdd que contenga datos de carácter personal (cualquier información numérica, alfabética, gráfica, fotográfica, acústica o de cualquier otro tipo concerniente a personas físicas identificadas o identificables.).

Por lo cual, el objetivo principal de este proyecto es:

- Establecer una métrica para medir la calidad de los datos que se van a introducir en una base de datos.

Para introducir al lector en el tema del proyecto, se desarrollarán las siguientes actividades:

- Realizar un amplio estudio del estado del arte de la calidad de los datos en la actualidad.
- Explicar el concepto de auditoría de calidad.

Hitos deseables:

- Destacar la importancia de la calidad de los datos.
- Provocar, al menos, al lector, la curiosidad de querer medir la calidad de los registros que desee introducir en su base de datos.

Objetivo aplicable a la parte práctica del proyecto:

- Se desea comprobar la calidad de los datos a introducir en distintos modelos de bases de datos. La finalidad de las diferentes bases de datos serán las siguientes: Universidad, Banco, Web de Empleo y Videoclub, tratándose todas ellas de bases de datos de carácter personal.  
Para medir la calidad con la que se introducen los datos, se construirá una métrica, a partir de la cual, el usuario que quiera saber el nivel de calidad de la base de datos, podrá ver un sencillo dato, como será un



tanto por ciento de calidad. Este tanto por ciento mostrará en un rango de 0 a 100 el nivel de calidad en los datos, donde el valor '0' indicará una falta total de calidad, y el valor '100' una perfecta calidad en los datos introducidos.

## 1.3 Fases del desarrollo

Para llevar a cabo el desarrollo este Proyecto de Fin de Carrera, se ha dividido en varias fases distintas:

- La primera, meramente teórica, centrada en la que la búsqueda de información del tema elegido, a través de las diferentes fuentes encontradas (libros, sitios Web,...), la cual fue muy importante, para así prever una primera estructura del proyecto a realizar.
- Una segunda etapa, en la que se estudió la problemática del tema elegido, además de la situación actual (estado del arte) además de todas las implicaciones legales que la calidad implica, y todas las posibles soluciones a aportar, para intentar mitigar, en lo posible, dicha problemática.
- Y, por último, una tercera etapa de contenido práctico, en la que se expuso la solución obtenida, además de mostrar las herramientas de apoyo al auditor, como son el servicio Web y el Checklist implementados.



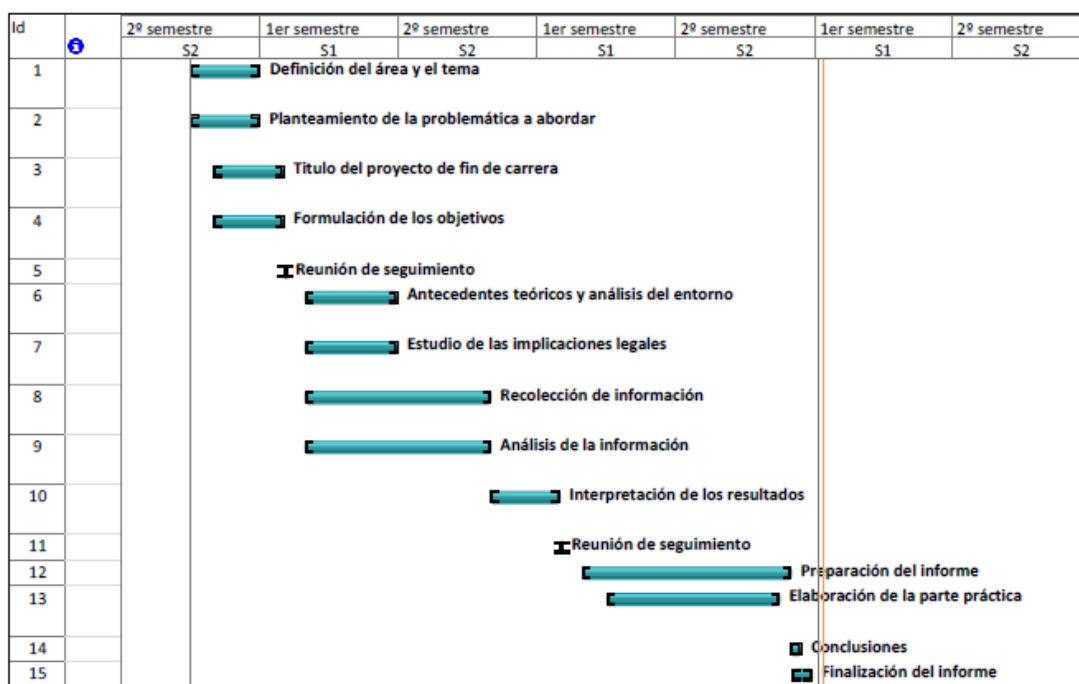


## Auditoría de calidad de bases de datos de carácter personal

A continuación se muestra un calendario del desarrollo del proyecto (dividido por semestres), incluyendo la duración de cada una de las tareas efectuadas durante la elaboración de este Proyecto Fin de Carrera a modo de resumen:

Nombre de tarea	Duración
Definición del área y el tema	1 ms
Planteamiento de la problemática a abordar	1 ms
Título del proyecto de fin de carrera	2 sem.
Formulación de los objetivos	2 sem.
Reunión de seguimiento	2 horas
Antecedentes teóricos y análisis del entorno	2 sem.
Estudio de las implicaciones legales	2 sem.
Recolección de información	2 mss
Análisis de la información	2 mss
Interpretación de los resultados	3 sem.
Reunión de seguimiento	2 horas
Preparación del informe	1 ms
Elaboración de la parte práctica	1,5 mss
Conclusiones	1 día
Finalización del informe	1 día
Reunión de seguimiento	2 horas

**Ilustración 1. Duración de cada actividad del proyecto**



**Ilustración 2. Calendario del desarrollo del proyecto**

Como se puede observar el desarrollo del proyecto empezó en el segundo semestre del año 2010 (octubre), acabando el mismo al final del segundo semestre del año 2012.



## 1.4 Medios empleados

Para la elaboración de este proyecto los medios empleados han sido mínimos. Se ha contado con un ordenador portátil del fabricante HP, además de software de Microsoft, como es el Word, el Excel y el Project. Adicionalmente a este software, se ha utilizado el Microsoft Visual Studio, en su versión 2010 (edición Ultimate), para la elaboración de la parte práctica del proyecto.

Para este proyecto no se ha requerido la presencia en ningún laboratorio de ningún tipo.

## 1.5 Estructura de la memoria

El proyecto se organiza en torno a la creación de una métrica, que, mediante la evaluación de la calidad de los datos introducidos en una base de datos de carácter personal, sirva de apoyo a las decisiones del auditor final. Este proyecto está dividido en dos partes bien diferenciadas, una parte teórica y una parte práctica.

- Parte teórica

Realizaremos un estudio para conocer el concepto de auditoría de la calidad en los datos. Una vez estudiado el concepto veremos cómo aplicarlo, en forma de métrica a una base de datos de carácter personal.

- Parte práctica

Con respecto a la parte práctica de este proyecto, realizaremos un servicio web que comprenderá una pequeña funcionalidad. Ésta contendrá un formulario en el que un usuario externo podrá introducir los campos relativos a sus datos personales, como pueden ser el nombre, sus apellidos o su domicilio fiscal, y el sistema le indicará el nivel de calidad de los datos que ha introducido.



A continuación, me dispongo a especificar la distribución de este Proyecto Fin de Carrera, resumiendo cada una de las partes que constituyen el mismo:

### ***Capítulo I – Introducción y objetivos del proyecto.***

En este capítulo comenzaré por explicar el contenido del proyecto de forma general y esquemática, con el fin de comprender de primera mano cómo evolucionará este.

### ***Capítulo II – Estado del arte.***

En este pequeño capítulo se introducirá al lector en el problema de la calidad de datos, la importancia que se le debe dar hoy en día y como evaluarlo.

### ***Capítulo III - Auditoría informática.***

En este capítulo se explica el concepto de auditoría informática. Estudiaremos el concepto de auditoría interna y externa, con las metodologías posibles y herramientas que se utilizan, además de dar una definición para poder diferenciar consultoría de auditoría.

### ***Capítulo IV - Bases de datos.***

En este capítulo se explica brevemente el concepto de base de datos, además del de sistema gestor de bases de datos.

### ***Capítulo V - Auditoría de calidad.***

En esta tercera parte explicaremos el concepto de auditoría de calidad. Nos centraremos en el concepto de calidad de los datos, sus beneficios y sus ventajas frente a las consecuencias que pueden llegar a tener. Explicaremos también el coste que puede llegar a tener la falta de calidad de datos en las empresas.

### ***Capítulo VI - Implicaciones legales.***

Llegados a este punto, en este capítulo mostraremos el capítulo de la Ley Orgánica de Protección de Datos de carácter personal referente a la calidad de los datos, además del ámbito de aplicación y los antecedentes normativos. Podremos ver un ejemplo de cómo una persona física puede dar de alta un



Auditoría de calidad de bases de datos de carácter personal  
fichero en la Agencia Española de Protección de Datos, y mostraremos un  
historial de operaciones en los Registros Generales de Protección de Datos.

### ***Capítulo VII – Soluciones comerciales***

En este capítulo, una vez introducido al lector en todo el contexto teórico, se mostrarán las soluciones comerciales que existen en la actualidad en referencia al proyecto tratado, contando sus ventajas e inconvenientes encontrados.

### ***Capítulo VIII - Métrica.***

En esta parte se definirá la métrica a utilizar para medir el nivel de calidad o de salud de los datos a introducir en las diferentes bases de datos definidas.

### ***Capítulo IX - Herramienta de apoyo al auditor.***

En este capítulo se propone una herramienta de apoyo al auditor en la realización de su tarea para medir el nivel de calidad de una base de datos, además del uso de un checklist.

### ***Capítulo X - Presupuesto.***

En dicho capítulo se detallará la valoración económica del Proyecto realizado.

### ***Capítulo XI - Líneas de investigación futuras.***

En este capítulo se exponen las diferentes líneas de investigación futuras para proyectos futuros sobre temas de la calidad en los datos.

### ***Capítulo XII - Conclusiones.***

Una vez llegado a este capítulo, se expondrán las conclusiones sobre el proyecto, basadas en documentación y los conocimientos adquiridos durante la elaboración de este Proyecto.

### ***- Anexo.***

- Bibliografía.

Amplia documentación de las referencias y sitios Web utilizados para la exposición y argumentación del Proyecto.

- Glosario de términos y acrónimos.



Diccionario técnico en el que se almacenarán todos los términos y acrónimos utilizados en el documento del Proyecto.

- Formulario NOTA en la AEPD.

¿Cómo se da de alta una base de datos, o un fichero en la AEPD? Quiénes deben notificar, pasos a seguir para cumplimentar el fichero y resumen de las estadísticas mensuales en el RGPD.

- Resumen resoluciones LOPD.

Resumen de las resoluciones citadas de la Ley Orgánica de Protección de Datos en el Proyecto.

- Ponderaciones de los campos de las diferentes bases de datos.

Breve explicación del motivo de las ponderaciones de los campos en las diferentes bases de datos que se explican con anterioridad.

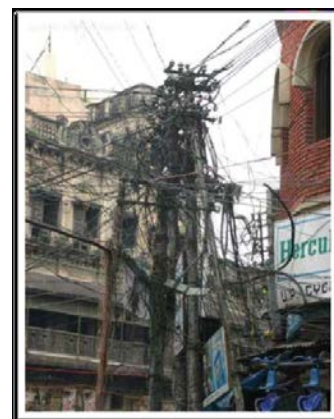
# Bloque II

## Estado del arte

Las bases de datos son el centro de atención para las grandes empresas de hoy en día, ya que constituyen uno de los soportes fundamentales para el proceso de toma de decisiones gerenciales; de ahí la importancia de que los datos guardados en ellas sean confiables y de calidad. Uno de los procesos en la construcción de estas y que contribuye a lograr este objetivo, es la limpieza de los datos.

Las empresas han ido tomando consciencia de la importancia de la Calidad de sus Datos para ahorrar costes y hacer rentables las inversiones realizadas en los complejos sistemas de CRM. Empiezan a darse cuenta de que la calidad de los datos no es solo responsabilidad del área de Tecnología, sino de toda la empresa, puesto que es un asunto para situar en el mismo nivel que cualquier otro servicio crítico de la empresa.

Pero, ¿por qué limpiar las bases de datos? Actualmente no disponemos de una fuente única y confiable de la verdad, no tenemos capacidad de mantenimiento de las dimensiones, obtenemos ideas contradictorias de los productos, clientes, etc., no existe un sistema para mantener los atributos, existen múltiples ubicaciones del mismo registro, y lo peor de todo, estamos rodeados de inconsistencias de costes, errores, informes incorrectos, etc.



**Ilustración 3. Bases de datos sin limpiar**



**Ilustración 4. Costes de la mala calidad**

Alejandra Viana Lara

Los costes asociados con mala calidad en los datos incluyen, entre otras cosas, un tiempo invertido en negocio innecesario, una comunicación inapropiada con los clientes, además de una información no fiable en la toma de decisiones.



Debido a esto el mundo empresarial ya cree que la calidad no es un punto de competitividad sino que es un procedimiento estándar.

Podemos resumir que la calidad de los datos se ha convertido en el centro de los esfuerzos de gestión de información en muchas organizaciones de la actualidad. Hablamos de los procesos de medición de la precisión, integridad y coherencia de los datos. Los problemas de calidad de datos se extienden como una plaga por empresas y organismos públicos de todo tipo y tamaño debido sobre todo al vacío normativo existente en la actualidad (incluso a nivel mundial), pues existen muy pocas normas en las que las empresas puedan basarse para mantener una buena calidad de datos. Existen infinidad de normas, como la ISO: 25000, la ISO: 12207, la IEEE 730 o la ISO / IEC 15504, etc., que tratan la calidad de los procesos que tratan los datos, y no tratan la calidad de datos en sí.

Cabe recalcar que la calidad y confiabilidad de los datos con los que contamos es directamente proporcional al éxito en la gestión de los objetivos y metas y toma de decisiones de toda empresa.

Por todo ello es vital para una empresa contar con información exacta o lo más confiable posible. Por lo cual, el canal de introducción de la información debe ser por las aplicaciones normalizadas para la introducción de esos datos.

El objetivo final, buscado por todas las empresas, en su lucha por mejorar la calidad de los datos que posee acerca de sus clientes, contratos, etc., consistirá en asegurarse una mayor integridad de la información, una mejora de la consistencia de la información, y, lo que es más importante, ofrecer una confianza de que los datos que poseen son exactos y están totalmente actualizados.



# Auditoría informática

## 3.1 Introducción

A finales del siglo XX, los Sistemas Informáticos se han constituido en las herramientas más poderosas para materializar uno de los conceptos más vitales y necesarios para cualquier organización empresarial, los Sistemas de Información de la empresa.

La Informática hoy, está subsumida en la gestión integral de la empresa, y por eso las normas y estándares propiamente informáticos deben estar, por lo tanto, sometidos a los generales de la misma. En consecuencia, las organizaciones informáticas forman parte de lo que se ha denominado el "management" o gestión de la empresa. Cabe aclarar que la Informática no gestiona propiamente la empresa, ayuda a la toma de decisiones, pero no decide por sí misma. Por ende, debido a su importancia en el funcionamiento de una empresa, existe la Auditoría Informática.

El término de Auditoría se ha empleado incorrectamente con frecuencia ya que se ha considerado como una evaluación cuyo único fin es detectar errores y señalar fallos. A causa de esto, se ha tomado la frase "Tiene Auditoría" como sinónimo de que, en dicha entidad, antes de realizarse la auditoría, ya se habían detectado fallos.

El concepto de auditoría es mucho más que esto. **Es un examen crítico que se realiza con el fin de evaluar la eficacia y eficiencia de una sección, un organismo, una entidad, etc.**

La palabra auditoría proviene del latín *auditorius*, y de esta proviene la palabra auditor, que se refiere a todo aquel que tiene la virtud de oír.

Por otra parte, el diccionario Español Sopena lo define como: Revisor de Cuentas colegiado. En un principio esta definición carece de la explicación del objetivo fundamental que persigue todo auditor: evaluar la eficiencia y eficacia.

De todo esto sacamos como deducción que la auditoría es un examen crítico pero no mecánico, que no implica la preexistencia de fallos en la entidad auditada y que persigue el fin de evaluar y mejorar la eficacia y eficiencia de una sección o de un organismo.





Aunque, se podría definir también, como **el proceso de recoger, agrupar y evaluar evidencias para determinar si un Sistema de Información salvaguarda el activo empresarial, mantiene la integridad de los datos, lleva a cabo eficazmente los fines de la organización y utiliza eficientemente los recursos.**

Los principales objetivos que constituyen a la auditoría Informática son el control de la función informática, el análisis de la eficiencia de los Sistemas Informáticos que comporta, la verificación del cumplimiento de la Normativa general de la empresa en este ámbito y la revisión de la eficaz gestión de los recursos materiales y humanos informáticos.

El auditor informático ha de velar por la correcta utilización de los amplios recursos que la empresa pone en juego para disponer de un eficiente y eficaz Sistema de Información. Claro está, que para la realización de una auditoría informática eficaz, se debe entender a la empresa en su más amplio sentido, ya que una Universidad, un Ministerio o un Hospital son tan empresas como una Sociedad Anónima o empresa Pública. Todos utilizan la informática para gestionar sus "negocios" de forma rápida y eficiente con el fin de obtener beneficios económicos y de costes.



### **3.2 Auditoría informática**

La auditoría nace como un órgano de control de algunas instituciones estatales y privadas. Su función inicial es estrictamente económico-financiera, y los casos inmediatos se encuentran en las peritaciones judiciales y las contrataciones de contables expertos por parte de Bancos Oficiales.

La función auditora debe ser absolutamente independiente; no tiene carácter ejecutivo, ni son vinculantes sus conclusiones. Queda a cargo de la empresa tomar las decisiones pertinentes. La auditoría contiene elementos de análisis, de verificación y de exposición de debilidades y disfunciones. Aunque pueden aparecer sugerencias y planes de acción para eliminar las disfunciones y debilidades antedichas; estas sugerencias plasmadas en el Informe final reciben el nombre de Recomendaciones.

Las funciones de análisis y revisión que el auditor informático realiza, puede chocar con la psicología del auditado, ya que es un informático y tiene la necesidad de realizar sus tareas con racionalidad y eficiencia. La reticencia del auditado es comprensible y, en ocasiones, fundada. El nivel técnico del auditor es a veces insuficiente, dada la gran complejidad de los Sistemas, unidos a los plazos demasiado breves de los que suelen disponer para realizar su tarea.

Además del chequeo de los Sistemas, el auditor somete al auditado a una serie de cuestionario. Dichos cuestionarios, llamados Check List, son guardados celosamente por las empresas auditoras, ya que son activos importantes de su actividad. Las Check List tienen que ser comprendidas por el auditor al pie de la letra, ya que si son mal aplicadas y mal recitadas se pueden llegar a obtener resultados distintos a los esperados por la empresa auditora. La Check List puede llegar a explicar cómo ocurren los hechos pero no por qué ocurren. El cuestionario debe estar subordinado a la regla, a la norma, al método. Sólo una metodología precisa puede desentrañar las causas por las cuales se realizan actividades teóricamente inadecuadas o se omiten otras correctas.

El auditor sólo puede emitir un juicio global o parcial basado en hechos y situaciones incontrovertibles, careciendo de poder para modificar la situación analizada por él mismo.

*La Auditoría Informática la podemos definir como “el conjunto de procedimientos y técnicas para evaluar y controlar un sistema informático con el fin de constatar si sus actividades son correctas y de acuerdo a las normativas informáticas y generales prefijadas en la organización”.*



### **3.3 Auditoría interna y auditoría externa**

Existen dos tipos de auditoría: interna y externa.

La *auditoría interna* es la realizada con recursos materiales y personas que pertenecen a la empresa auditada. Los empleados que realizan esta tarea son remunerados económicamente. La auditoría interna existe por expresa decisión de la Empresa, o sea, que puede optar por su disolución en cualquier momento.

Este tipo de auditoría, revisa cada departamento con cierta periodicidad. Su acceso ha de ser suficiente para poder realizar informes objetivos e independientes que pueden hacer llegar a quienes no estén implicados en los procesos auditados.

Jerárquicamente pueden depender de Dirección general y auditor general, aunque sería responsable un comité de auditoría. La elección del personal sería:

- De la entidad: se valora formación y experiencia, salvaguardar hándicap, y objetividad con el compañero.
- Externo: se valora la experiencia afín y el conocimiento en el sector.

Por otro lado, la *auditoría externa* es realizada por personas afines a la empresa auditada; es siempre remunerada. Se presupone una mayor objetividad que en la Auditoría Interna, debido al mayor distanciamiento entre auditores y auditados.

Realiza revisiones anuales (aunque varía dependiendo del contrato), y debe hacer una contratación formal que especifique el alcance del proceso de auditoría y el proceso geográfico.

La contratan los directores de administración, el consejo de la empresa, accionistas...

Por último, en este tipo de auditoría, es recomendable establecer un plan de trabajo a realizar. Son necesarias unas cláusulas de confidencialidad.

*Relación de los dos tipos de auditoría:*

La auditoría interna y la auditoría externa son compatibles y recomendables, además de complementarias.

Las externas pueden apoyarse en las internas sin perder su objetividad (informes) y aportar técnicas y métodos. Por otro lado, las internas pueden crearse por recomendación de las externas y en los informes de estos puede haber denuncias hacia los internos.

Entre auditores internos y externos existirá un trato fluido. Ambos tendrán relación con otras áreas de la entidad (en función del objeto de la auditoría).



### **3.4 Alcance de la auditoría informática**

El alcance ha de definir con precisión el entorno y los límites en que va a desarrollarse la auditoría informática. El alcance ha de figurar expresamente en el Informe Final, de modo que quede perfectamente determinado no solamente hasta que puntos se ha llegado, sino cuales materias fronterizas han sido omitidos. En este sentido un ejemplo de este control surge al plantearse las siguientes cuestiones ¿Se someterán los registros grabados a un control de integridad exhaustivo? ¿Se comprobará que los controles de validación de errores son adecuados y suficientes?

En definitiva, el alcance de la auditoría, concepto de vital importancia, nos viene dado por los procedimientos. La amplitud y profundidad de los procedimientos que se apliquen nos definen su alcance.

Se pretende garantizar que se toman en consideración todos los aspectos, áreas, elementos, operaciones, circunstancias, etc. Que sean significativas.

Para ello se establecen unas normas y procedimientos que en cuanto a la ejecución de la auditoría se resumen en que:

- El trabajo se planificará apropiadamente y se supervisará adecuadamente.
- Se estudiará y evaluará el sistema de control interno.
- Se obtendrá evidencia suficiente y adecuada.

Como corolario se establece que la evidencia obtenida deberá recogerse en los papeles de trabajo del auditor como justificación y soporte del trabajo efectuado y la opinión expresada.

Por último, cabe recalcar, que la indefinición de los alcances de la auditoría puede comprometer el éxito de la misma.



### **3.5 Síntomas de necesidad de una auditoría informática**

Las empresas acuden a las auditorías externas cuando existen síntomas bien perceptibles de debilidad. Estos síntomas pueden agruparse en clases:

- Síntomas de descoordinación y desorganización:
  - No coinciden los objetivos de la Informática de la Compañía y de la propia Compañía.
  - Los estándares de productividad se desvían sensiblemente de los promedios conseguidos habitualmente.
- Síntomas de mala imagen e insatisfacción de los usuarios:
  - No se atienden las peticiones de cambios de los usuarios. Ejemplos: cambios de Software en los terminales de usuario, refrescamiento de paneles, variación de los ficheros que deben ponerse diariamente a su disposición, etc.
  - No se reparan las averías de Hardware ni se resuelven incidencias en plazos razonables. El usuario percibe que está abandonado y desatendido permanentemente.
  - No se cumplen en todos los casos los plazos de entrega de resultados periódicos. Pequeñas desviaciones pueden causar importantes desajustes en la actividad del usuario, en especial en los resultados de Aplicaciones críticas y sensibles.
- Síntomas de debilidades económico-financiero:
  - Incremento desmesurado de costes.
  - Necesidad de justificación de Inversiones Informáticas (la empresa no está absolutamente convencida de tal necesidad y decide contrastar opiniones).
  - Desviaciones Presupuestarias significativas.
  - Costes y plazos de nuevos proyectos (deben auditarse simultáneamente a Desarrollo de Proyectos y al órgano que realizó la petición).
- Síntomas de Inseguridad: Evaluación de nivel de riesgos
  - Seguridad Lógica
  - Seguridad Física
  - Confidencialidad
  - Continuidad del Servicio. Es un concepto aún más importante que la Seguridad. Establece las estrategias de continuidad entre fallos mediante Planes de Contingencia\* Totales y Locales.



### Auditoría de calidad de bases de datos de carácter personal

- Centro de Proceso de Datos fuera de control. Si tal situación llegara a percibirse, sería prácticamente inútil la auditoría. Esa es la razón por la cual, en este caso, el síntoma debe ser sustituido por el mínimo indicio.

Debido a estos factores se establecen unos **planes de contingencia**, que son un tipo de plan preventivo, predictivo y reactivo. Presentan una estructura estratégica y operativa que ayudará a controlar una situación de emergencia y a minimizar sus consecuencias negativas.

Por ejemplo, la empresa sufre un corte total de energía o explota, ¿Cómo sigo operando en otro lugar? Lo que generalmente se pide es que se hagan Backups de la información diariamente y que aparte, sea doble, para tener un Backup en la empresa y otro afuera de ésta. Una empresa puede tener unas oficinas paralelas que posean servicios básicos (luz, teléfono, agua) distintos de los de la empresa principal. En este caso, si se produce la inoperancia de Sistemas en la empresa principal, se utilizaría el Backup para seguir operando en las oficinas paralelas. Los Backups se pueden acumular durante dos meses, o el tiempo que estipule la empresa, y después se van reciclando.



### 3.6 Herramientas y técnicas para la auditoría informática

En la auditoría informática tenemos, por orden de mayor importancia y generalidad, las normas, los procedimientos, las técnicas y las herramientas. El auditor, en el desarrollo de su labor, ha de realizar su examen conforme a las normas técnicas, oficiales o generalmente reconocidas, como reglas técnicas y éticas. Según la naturaleza del sistema o de la auditoría, el auditor tiene a su disposición una serie de técnicas, que le ofrecen diversas alternativas contrastadas en la evaluación de elementos específicos del sistema a auditar. Sin embargo, el auditor, solamente con su intelecto y sus manos, no puede llegar muy lejos en la obtención y elaboración de la información, y sobre todo cuando el entorno es muy especializado. Entonces, le hace falta un conjunto de herramientas que le facilite su labor y le ayude a obtener datos con mayor precisión y fiabilidad.

Una herramienta de auditoría se puede definir como el elemento que, mediante la estructuración, clasificación, y automatización de determinados procedimientos de auditoría, facilita el trabajo del auditor al descargarle de tareas rutinarias, repetitivas o extremadamente exhaustivas, y permitirle centrarse en aspectos cualitativos o fundamentales.

Las herramientas de auditoría pueden ser de muchos tipos y presentarse en diversos soportes materiales. Estas herramientas, se pueden dividir de acuerdo a dos criterios principales, en cuanto a la innovación tecnológica y en cuanto a su especialización.

1) En cuanto a innovación tecnológica:

- Herramientas tradicionales: son las herramientas que ha venido utilizando el auditor normalmente. Estas son:
  - Cuestionarios previos
  - Estándares
  - Entrevistas
  - **Checklist**
  - Matrices de riesgos.
- Herramientas informáticas: son las que funcionan propiamente con el ordenador:
  - Trazas y/o huellas de auditoría.
  - Software estadístico.
  - Software de control de proyectos.
  - Software de interrogación.
  - Software de ofimática (Procesadores de texto, hojas de cálculo).
  - Lenguaje de programación.
  - Simuladores y analizadores de ssos.
  - Monitores.
  - Software de re-ingeniería e ingeniería inversa.
  - Paquetes de auditoría (CAAT).



2) En cuanto a su especialización

- Herramientas de propósito general: son las herramientas cuyo espectro de aplicación trasciende del ámbito de la auditoría:
  - Estándares.
  - Software de ofimática.
  - Software de control de proyectos.
  - Software de interrogación.
  - Lenguajes de programación.
  - Simuladores y analizadores de ssos.
  - Monitores.
  - Software de re-ingeniería e ingeniería inversa.
- Herramientas de propósito específico: son las que tienen solamente aplicación en la actividad informática:
  - Cuestionarios previos.
  - Entrevistas.
  - **Checklist.**
  - Trazas y/o huellas de auditoría.
  - Paquetes de auditoría (CAAT).





### **3.7 Metodología de trabajo de auditoría informática**

El método de trabajo del auditor pasa por las siguientes etapas:

- *Alcance y Objetivos de la Auditoría Informática.*
- *Estudio inicial del entorno auditable.*
- *Determinación de los recursos necesarios para realizar la auditoría.*
- *Elaboración del plan y de los Programas de Trabajo.*
- *Actividades propiamente dichas de la auditoría.*
- *Confección y redacción del Informe Final.*
- *Redacción de la Carta de Introducción o Carta de Presentación del Informe final.*

### **3.8 Auditoría VS consultoría**

Llegados a este punto, es necesario aclarar cuáles son las diferencias de la auditoría con el concepto de consultoría.

El matiz principal está en el objeto del trabajo y la profundidad del informe, además de la implicación en el proceso de selección e implantación de los sistemas.

El auditor hace una evolución profunda (analiza, prueba...) y lo plasma en un informe. Por el contrario, el consultor pone en marcha el sistema de solución encontrado, implementa la recomendación. Busca soluciones a problemas y los resuelve.

Además, la consultoría es una función a priori con el fin de determinar cómo llevar a cabo una función o actividad de forma que obtenga los resultados pretendidos. La auditoría, sin embargo, verifica a posteriori si estas condiciones, una vez realizada esta función o actividad, se cumplen y los resultados pretendidos se obtienen realmente.

Para concluir, diremos que es recomendable auditar y luego consultoría, y que sean diferentes personas quienes lo lleven a cabo, pues si no, luego no se podrá realizar una auditoría de S.I.



# Bases de datos

## 4.1 Introducción

La cantidad de información disponible crece, literalmente, de manera explosiva, y el valor de los datos como activo de las organizaciones está ampliamente reconocido. Para que los usuarios obtengan el máximo rendimiento de su enormes y complejos conjuntos de datos son necesarias herramientas que simplifiquen las tareas de administrar los datos y de extraer información útil en el momento preciso. En caso contrario, los datos pueden convertirse en una carga cuyo coste de adquisición y de gestión supere ampliamente el valor obtenido de ellos.

Las bases de datos y los sistemas de bases de datos se han convertido en elementos imprescindibles en la vida cotidiana de la sociedad moderna. Cada día, la mayoría de nosotros nos encontramos con actividades que requieren algún tipo de interacción con una base de datos.

## 4.2 Bases de datos

Una base de datos es un conjunto de datos, que generalmente describe las actividades de una o varias organizaciones relacionadas.

Por datos, queremos decir hechos conocidos que pueden registrarse y que tienen un significado implícito.

En otras palabras, una base de datos tiene alguna fuente de la cual provienen los datos, algún grado de interacción con los sucesos del mundo real, y una audiencia que está activamente interesada en el contenido de la base de datos.



Una base de datos puede tener cualquier tamaño y complejidad, y puede crearse y mantenerse manualmente o puede ser informatizada mediante un conjunto de programas de aplicación diseñados específicamente para dicha tarea o bien mediante un sistema de gestión de bases de datos.

### **4.3 Sistemas gestores de bases de datos**

Un sistema de gestión de bases de datos (database management system) es una colección de programas que permiten a los usuarios crear y mantener una base de datos. El SGBD es por tanto un sistema de software de propósito general que facilita los procesos de definición, construcción y manipulación de bases de datos para distintas aplicaciones.

La definición de una base de datos consiste en especificar los tipos de datos, las estructuras y restricciones para los datos que se van a almacenar en dicha base.

La construcción de la base de datos es el proceso de almacenar los datos concretos sobre algún medio de almacenamiento controlado por el SGBD.

La manipulación de la base de datos incluye funciones tales como consultar la base de datos para recuperar unos datos específicos, actualizar la base de datos para reflejar los cambios ocurridos en el Universo del Discurso, y generar informes a partir de ellos.

No es preciso utilizar software de SGBD de propósito general para implementar una base de datos informatizada. Podríamos codificar nuestro propio conjunto de programas para crear y mantener la base de datos, es decir, usemos o no un SGBD de propósito general, normalmente tendremos que emplear gran cantidad de software para manipular la base de datos.

A continuación citaremos algunas de las ventajas de utilizar un SGBD:

- Control de la redundancia.
- Restricción de los accesos no autorizados.
- Suministro de almacenamiento persistente de objetos y estructuras de datos de programas.
- Suministro de copias de seguridad y recuperación.
- Suministro de múltiples interfaces de usuario.
- Independencia con respecto a los datos.
- Acceso eficiente a los datos.
- Reducción del tiempo de desarrollo de las aplicaciones.

Aunque, a pesar de las ventajas que supone utilizar un SGBD, hay algunas situaciones en las que el empleo de un SGBD puede generar costes adicionales innecesarios:

- Una fuerte inversión inicial en equipo, software y capacitación.
- La generalidad que ofrece el SGBD para definir y procesar datos.



- Los costos que implica ofrecer las funciones de seguridad, control de concurrencia, recuperación e integridad.

# Auditoría de la calidad

## 5.1 ¿Qué es la calidad?

La palabra **calidad** tiene múltiples significados. Es un conjunto de propiedades inherentes a un objeto que le confieren capacidad para satisfacer necesidades implícitas o explícitas. La calidad de un producto o servicio es la percepción que el cliente tiene del mismo, es una fijación mental del consumidor que asume conformidad con dicho producto o servicio y la capacidad del mismo para satisfacer sus necesidades.

Según la RAE la calidad se define como la “Propiedad o conjunto de propiedades inherentes a algo, que permiten juzgar su valor”.

La calidad ha dejado de ser un tópico, y forma parte de los productos o servicios que comercializamos para nuestros clientes. Cada vez exigimos más que los productos o servicios que nos suministran nuestros proveedores tengan el mayor grado de calidad dentro de un precio razonable.

El cliente es el mejor auditor de la calidad, él exige el nivel que está dispuesto a pagar por ella. Por tanto, debemos de cuantificar cual es el nivel de calidad que nos exige para poder planificar la calidad de los productos semi-elaborados que se generen a lo largo del proceso de producción del producto o servicio final.

Al analizar las necesidades de nuestros clientes, deberemos tener en cuenta también la previsible evolución de sus necesidades y tendencias en cuanto a características. Deberemos tener en cuenta la evolución tecnológica del entorno de producción de nuestros productos para suministrarles con el nivel tecnológico adecuado. No debemos olvidar tampoco el nivel de calidad de nuestros competidores, debiendo elaborar productos cuyas características y funcionalidades sean competitivas con las de nuestros competidores, así como su calidad.

La calidad se ha convertido en el medio de subsistir dentro de un mercado competitivo, lo cual beneficia al consumidor final, es decir, a nosotros. Es el primer filtro lógico por el que las



### Auditoría de calidad de bases de datos de carácter personal

empresas prevalecen en el mercado, el segundo será la productividad que emplean para conseguir esa calidad.

La calidad será el objetivo global a conseguir, y la productividad nos vendrá por añadidura, nunca al revés.

¿Entonces, a partir de aquí, como definimos la Auditoría de Calidad?

La norma ISO 9000:2000 (predecesora de la norma ISO 9001: 2008) definía una Auditoría de Calidad como:

**“Proceso sistemático, independiente y documentado para obtener evidencias y evaluarlas de manera objetiva con el fin de determinar el alcance al que se cumplen los criterios de auditoría”.**



**Ilustración 5. ISO 9001:2008**



## **5.2 Objetivo de la auditoría de la calidad**

Una auditoría de Calidad tiene como objetivo el mostrar la situación real para aportar confianza y destacar las áreas que puedan afectar adversamente esa confianza.

El objetivo de la auditoría es evaluar la suficiencia y efectividad de las disposiciones de calidad de una organización mediante la recolección y uso de evidencia objetiva, e identificar y registrar las instancias de no cumplimiento con las disposiciones de calidad e indicar, donde sea posible, las razones.

Existen varias razones para realizar una auditoría:

- Establecer el estado de un proyecto.
- Verificar la capacidad de realizar o continuar un trabajo específico.
- Verificar qué elementos aplicables del programa o Plan de Aseguramiento de la Calidad han sido desarrollados y documentados.
- Verificar la adherencia de esos elementos con el programa o Plan de Aseguramiento de la Calidad.

El propósito y la actividad de la auditoría es recoger, examinar y analizar la información necesaria para tomar las decisiones de aprobación.

La auditoría debe tener capacidad para investigar la pericia técnica, el desarrollo del software o la capacidad del departamento de desarrollo, el esfuerzo disponible, el soporte del mantenimiento o la efectividad de la gestión.

En las auditorías debe acordarse el dirigirse a criterios específicos tales como la realización del código software.

Cuando se verifiquen los puntos débiles, los auditores deberán tomar una actitud positiva y utilizar sus conocimientos y experiencia para hacer recomendaciones constructivas. En realidad, una función del auditor es pactar la idoneidad de cualquier acción correctiva propuesta. Este papel, si es usado adecuadamente, es uno de los vínculos más valorados entre las partes.



### **5.3 Auditoría interna y externa de calidad**

La calidad es una ventaja competitiva para aquellas empresas que se han identificado con ella y en donde la gerencia se encuentra plenamente identificada con su alcance, repercusiones, con todo lo que ello puede generar.

Una vez que se ha inculcado obliga a la gerencia estar atenta sobre su comportamiento, cuidando, la forma como se está manteniendo, su alcance, logros, así como sus debilidades.

Por tanto, la Auditoria de la calidad es una importante actividad que permite a las organizaciones mejorar sus funciones en forma continua. La auditoría puede ser tanto interna o externa a la organización.

Auditoría Interna: Tiene como propósito verificar si las actividades relacionadas con la calidad están conformes con las especificadas en el sistema de asegurados y control de la calidad estas actividades las desarrolla personal de la empresa.

Las acciones conectivas después de una auditoría interna deben documentarse.

Auditoría Externa: Tiene el mismo propósito de la auditoría interna pero con la diferencia que es realizada por miembros que no pertenecen a la organización.

### **5.4 Repercusiones de la auditoría de calidad**

A continuación se citan algunas de las repercusiones más importantes que posee una auditoría de calidad:

- La auditoría de calidad permite al auditor, evaluar el grado de cumplimiento de los requisitos propuestos y, si es necesario introducir acciones de mejora; no se trata tanto de detectar fallos como de determinar la eficacia de las actuaciones para alcanzar los objetivos propuestos.
- Los resultados del informe (que son el producto final de la auditoria de calidad) proporcionan a la dirección de la empresa evidencias objetivas basadas en hechos. Estos permiten a la dirección tomar decisiones basándose en hechos y no en hipótesis.
- La auditoría de calidad tiene importantes repercusiones en toda la empresa ya que afecta a todas las funciones de la misma.



## **5.5 Ventajas de la auditoría de calidad**

Las principales ventajas de realizar una auditoría de calidad son las siguientes:

- La auditoría de calidad ayuda a verificar y evaluar las actividades relacionadas con la calidad en el seno de una organización, así como para la identificación y mejora de las áreas no conformes con el modelo exigido.
- Proporciona a la organización el poder ofrecer producto / servicio que satisfaga las necesidades de los clientes.
- Ser competitiva con respecto a otras empresas que ofrezcan el mismo producto / servicio.
- Poder ofrecer sus productos / servicios a nivel mundial y reconocida por su prestigio y buena calidad.
- Es una herramienta que contribuye al mejoramiento de la calidad dirigido al usuario que requiere de los producto / servicio.
- La empresa auditada mantiene al día sus actividades (procesos, producción, sistema etc.)

(Mora, 2009, <http://www.articuloz.com/empresas-articulos/la-importancia-de-la-auditoria-de-la-calidad-771130.html>)





## **5.6 Beneficios de la calidad de datos**

Calidad de datos se refiere a los procesos, técnicas, algoritmos y operaciones encaminados a mejorar la calidad de los datos existentes en empresas y organismos.

Sin embargo, la calidad de datos generalmente se refiere al mejoramiento de la calidad de los datos de personas físicas y jurídicas, pues son éstos probablemente los datos que más tienden a degradarse y cuya falta de calidad más impacta en la productividad de las organizaciones.

Los principales beneficios de la calidad de datos con algunos ejemplos incorporados son:

- Ahorrar costes directos: evitando tener información duplicada y por lo tanto evitar el envío replicado de cartas a un mismo cliente.
- Potenciar las acciones de marketing y la gestión: la normalización de ficheros mejora el análisis de datos y permite segmentaciones precisas para que sus acciones de marketing y su gestión ganen en precisión y eficacia.
- Optimizar la captación y la fidelización de clientes: con los datos correctos, se mejoran los ratios de respuestas y el cliente se siente plenamente identificado con la empresa.
- Mejorar la imagen corporativa: el cliente sólo recibe el envío que le corresponde, una sola vez y con sus datos correctos.
- Mejorar el servicio: identificación más rápidamente del cliente que llama a un Call Center, reduciendo los tiempos de espera y, dejando tiempo al operador para centrarse en el mensaje de negocio.



## 5.7 Problemas que afectan a la calidad de datos

Entre los problemas que afectan a la calidad de datos, están los *procesos externos*, los *procesos internos* y el *deterioro natural*.

Dentro de los procesos externos, podemos encontrarnos:

- Migración de datos: el proceso requiere establecer una correspondencia entre la estructura original y la nueva estructura. En teoría el problema es muy trivial pero en la práctica presenta muchas dificultades debido a valores faltantes o nulos, o a las condiciones específicas incorporadas en el código del programa. Con frecuencia hay duplicación de datos.
- Consolidación de sistemas: son parecidas a las migraciones, pero de mucha mayor complejidad. Usualmente los datos de las fuentes se trasladan a una base de datos que ya contiene información, lo cual genera toda clase de conflictos de datos, como duplicados, series de tiempo, etc.
- Entrada manual de datos: Algunas de las principales causas de errores son la captura errada del valor, valores faltantes, valores por defecto o incluso falta de instrucciones adecuadas.
- Alimentación por lotes: Las razones de las principales causas de errores pueden ser debidas a que usualmente no se someten estos procesos a pruebas regresivas y aseguramiento de la calidad por la falta de tiempo o por la dificultad de hacerlo.
- Interfases en tiempo real: en la actualidad existe muchos sistemas que intercambian muchos datos con interfaces en tiempo real. Esto permite tener la información sincronizada y es de alto valor para la organización aunque no da tiempo para verificar que los datos sean correctos.

Dentro de los procesos internos producidos, podemos encontrarnos:

- Procesamiento de datos.
- Limpieza de datos (estos procesos incluyen la validación de datos y la corrección de los mismos para poder alcanzar datos de calidad).
- Depuración de datos.

Y, dentro del deterioro natural que se produce, nos encontramos:

- Nuevos usos de los datos.
- Cambios no registrados.
- Actualización de los sistemas.
- Pérdida de conocimiento (una ausencia temporal o permanente de los expertos en los datos conduce al uso inapropiado que afecta a la calidad de datos).
- Automatización de procesos.



## **5.8 A quién afecta la no calidad de los datos**

Si determinados campos que contienen datos inadecuados para los objetivos que perseguimos, contienen valores desconocidos o incoherentes, están duplicados, faltan o están incompletos, o son erróneos, podemos afirmar que carecen de calidad. Pero, ¿a quién afecta este aspecto?

Todos los departamentos de su negocio están afectados por la no calidad de los datos:

- El departamento de compras no dispone de un control total sobre los pedidos a proveedores, las compras se realizan de una forma totalmente manual y artesanal, sin ninguna influencia del histórico de compras del año anterior o atendiendo a criterios de marketing de clientes.
- El departamento de marketing no sabe cuántos clientes tenemos ni sus perfiles de compra, comportamiento cíclico o historial de quejas y sugerencias; miles de mailings están duplicados o devueltos por direcciones incorrectas.
- El departamento comercial no posee un historial "inteligente" de ventas a clientes, no conoce los márgenes comerciales de negociación y no posee las herramientas para la toma de pedidos online en cualquier momento y lugar.
- El gerente no conoce los principales parámetros de la empresa cuando lo desea; no posee información agregada de grandes clientes ni posee los medios para que esta información sea perfilada y compartida en la empresa.

Esto es sólo una parte, porque quienes más están afectados de esta situación son los verdaderos protagonistas de su negocio: sus clientes:

- Perciben retrasos en sus pedidos o errores en entregas.
- Reciben mailings duplicados.
- Reciben ofertas de productos que acaban de adquirir.
- Observan un trato totalmente indiferenciado después de tantos años como cliente.



## 5.9 Procesos de calidad

En el entorno económico actual, la característica más importante es la competitividad, lo que quiere decir que los precios a los que ofrezcamos nuestros productos a nuestros clientes deben ser iguales o más bajos que los de la competencia, pero con una calidad más alta. Para conseguirlo es necesario tener una estructura de costes adecuada y disponer de una estrategia de costes adecuada y disponer de una estrategia de Calidad que afecte a todas las áreas de la entidad u organismo.

Para satisfacer los requisitos de calidad es necesario conocer las necesidades del cliente. Éstas vienen dadas por estos tres parámetros:

- Calidad de los productos y servicios.
- Plazo de entrega adecuado.
- Coste dentro de los límites fijados.

El establecimiento de acuerdos de nivel de servicio y el cumplimiento de sus requerimientos de dará un determinado grado de satisfacción, que deberemos saber medir todo una vez pasado el período de estabilización del producto entregado.

Una de las principales características de los procesos de calidad es la repetitividad de los mismos. Todo proceso debe estar suficientemente definido como para que pueda ser repetido consiguiendo los mismos resultados cada vez que se realice el mismo proceso.

Una vez alcanzada esta repetitividad de los procesos y teniendo elementos para medir los atributos de los productos obtenidos, trataremos de ir refinando el modelo del proceso para reducir los defectos entregados (definiendo defecto como cualquier variación de una característica establecida que origina el incumplimiento de las necesidades del cliente con la consiguiente insatisfacción del mismo).

Como se ha indicado anteriormente, las revisiones y las auditorías pueden usarse para actividades de aseguramiento de la calidad, gestión de proyectos, gestión de la configuración o funciones de control singulares.

A continuación se muestra una tabla de los principales procesos para conseguir objetivos de calidad.

OBJETIVOS	PRINCIPALES PROCESOS QUE INCLUYE
Evaluación	Revisiones de Gestión, Revisiones Técnicas
Verificación	Inspecciones, Walkthrough
Validación	Pruebas
Conformidad, Confirmación	Auditoría

**Tabla 1. Principales procesos en los objetivos de calidad**



### Auditoría de calidad de bases de datos de carácter personal

Cualquier proceso estándar tiene unas condiciones como prerequisites; éstas son necesarias, aunque no son suficientes en sí mismas para que el proceso quede completado.

Prerequisites en los procesos de revisión.

El objetivo de una revisión de un elemento software es evaluar el software o el estado, del proyecto para identificar las discrepancias sobre los resultados planificados y recomendar mejoras cuando sea apropiado.

El objetivo de la auditoría del software es suministrar una evaluación objetiva de los productos y los procesos para corroborar la conformidad con los estándares, las líneas guía, las especificaciones y los procedimientos. Los siguientes requerimientos para conseguir este objetivo:

- Objetivo de la auditoría, criterios existentes (por ejemplo, contratistas, requerimientos, planes, especificaciones, estándares) en relación con los elementos software y los procesos que puedan ser evaluados.
- El personal d auditoría es seleccionado para promover los objetivos del grupo. Son independientes de cualquier responsabilidad directa para los productos y los procesos examinados y pueden provenir de una organización externa.
- El personal de auditoría debe tener la suficiente autoridad que le permita una adecuada gestión con el fin de realizar la auditoría.



## 5.10 Factores de la calidad

Los factores que influyen en una calidad de datos, a la hora del modelado son los siguientes (Batini, Ceri y Navathe):

**Compleción:** un esquema es completo cuando representa todas las características relevantes del dominio de aplicación:

- Respecto a los requisitos.
- Respecto al esquema.

**Corrección:** un esquema es correcto cuando utilizan de forma apropiada los conceptos del modelo E/R:

- Sintáctica.
- Semántica.

**Minimalidad:** un esquema es minimal cuando todo aspecto de los requisitos aparece sólo una vez en el esquema.

**Expresividad:** un esquema es expresivo cuando representa los requisitos de una manera natural, sin necesidad de explicaciones adicionales.

**Legibilidad:** es una propiedad del diagrama que representa al esquema. Un diagrama tiene buena legibilidad cuando respeta ciertos criterios estéticos que hacen al diagrama elegante.

**Auto-explicación:** un esquema es auto-explicativo cuando un gran número de propiedades puede representarse utilizando el propio modelo conceptual, sin otros formalismos.

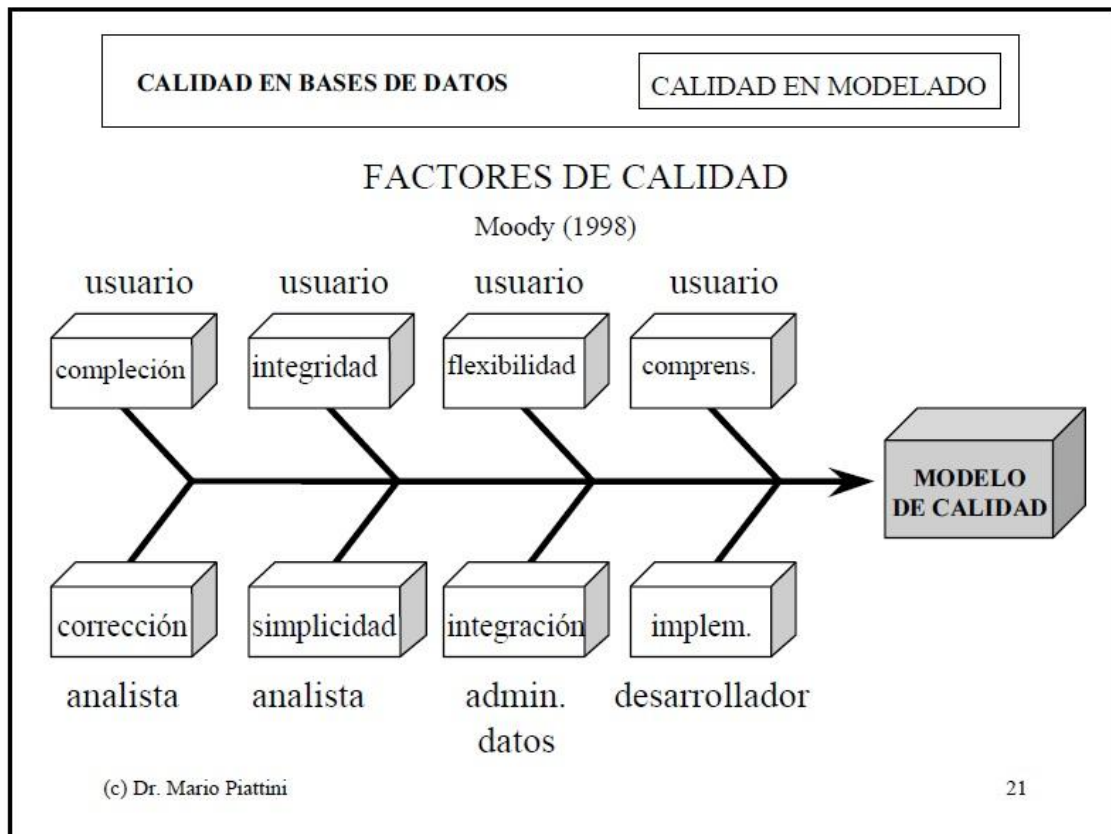
**Extensibilidad:** un esquema es fácilmente adaptado a cambios en los requisitos cuando se descompone en partes (módulos, vistas).

**Normalidad:** pretende conservar los datos en una forma limpia, “purificada”. (Aplicación de la teoría de la normalización del modelo relacional).



Ahora bien, ¿quién es la persona que se encarga de cada uno de estos factores en el proceso del modelado de una base de datos?

En el siguiente esquema, diseñado por Moody en el 1998, se detalla que perfil se encarga de cada uno de los factores de la calidad (según sus 8 factores):



**Ilustración 6. Factores de la calidad**

Como se puede observar, de los factores de completación, integridad, flexibilidad y comprensibilidad, se encarga el usuario, mientras que de la corrección y de la simplicidad se encarga el analista. De la integración se encarga el administrador de datos, y de la implementabilidad, por último, el desarrollador.

A continuación se explican cada uno de los ocho factores:

Completación:

- N° de elementos del modelo de datos que no corresponden con requisitos de usuario.
- N° de requisitos de usuario no representados en el modelo de datos.
- N° de elementos de datos que corresponden a requisitos de usuario pero definidos de forma inexacta.
- N° de inconsistencias con el modelo de procesos.



Integridad:

- N° de reglas del negocio que no se hacen cumplir por el modelo de datos.
- N° de restricciones de integridad incluidas en el modelo de datos que no corresponden a políticas del negocio.

Flexibilidad:

- N° de elementos en el modelo que están sujetos a cambios en el futuro.
- Costes estimados de los cambios.
- Importancia estratégica de los cambios.

Comprensibilidad:

- Valoración de los usuarios sobre la comprensibilidad del modelo.
- Capacidad de los usuarios de interpretar el modelo correctamente.
- Valoración de los desarrolladores de aplicaciones sobre la comprensibilidad del modelo.

Corrección:

- N° de violaciones de las convenciones de modelado de datos.
- N° de violaciones a las formas normales.
- N° de instancias de redundancia en el modelo.

Simplicidad:

- N° de entidades.
- N° de entidades e interrelaciones.
- N° de constructos ( $aNE + bNR + cNA$ ).

Integración:

- N° de conflictos con el modelo de datos corporativo.
- N° de conflictos con los sistemas existentes.
- Valoración de los representantes de todas las áreas de negocio.

Implementabilidad:

- Valoración de riesgo técnico.
- Valoración de riesgo de planificación.
- Estimación del coste de desarrollo.
- N° de elementos físicos incluidos en el modelo de datos.

(2012, <http://www.articuloz.com/empresas-articulos/la-importancia-de-la-auditoria-de-la-calidad-771130.html>)





### **5.11 Principio deontológico de la calidad**

Los principios deontológicos aplicables a los auditores deben necesariamente estar en consonancia con los del resto de profesionales y especialmente con los de aquellos cuya actividad presente mayores concomitancias con la de la auditoría.

El principio de calidad dice lo siguiente:

*El auditor deberá prestar sus servicios a tenor de las posibilidades de la ciencia y a medios a su alcance con absoluta libertad respecto a la utilización de dichos medios y en unas condiciones técnicas adecuadas para el idóneo cumplimiento de su labor.*

*En los casos en que la precariedad de medios puestos a su disposición impida o dificulten seriamente la realización de la auditoría, deberá negarse a realizarla hasta que se le garantice un mínimo de condiciones técnicas que no comprometan la calidad de sus servicios o dictámenes.*

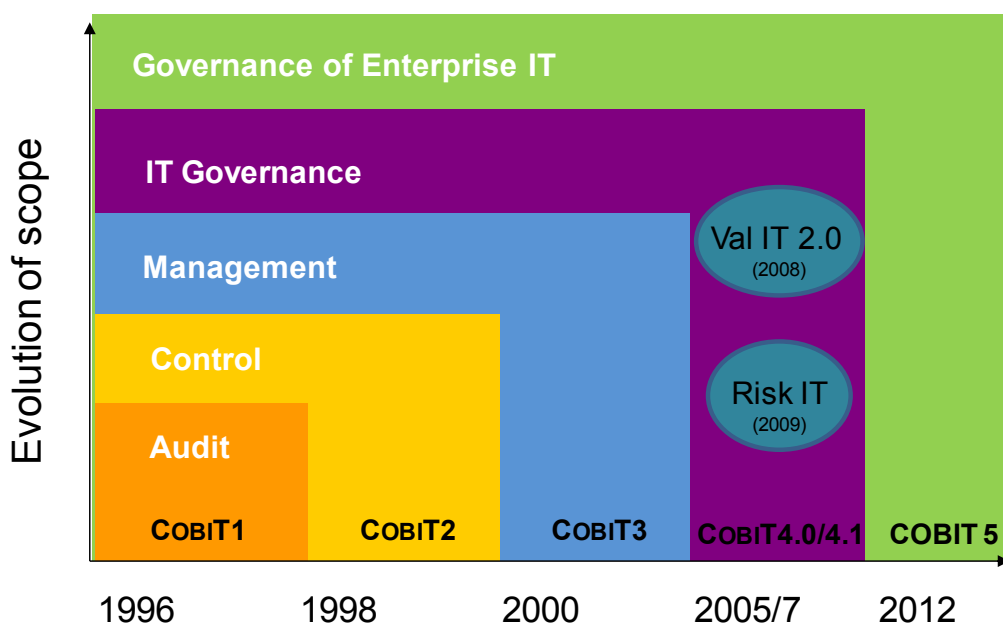
*Cuando durante la ejecución de la auditoría, el auditor considerase conveniente recabar el informa de otros técnicos más cualificados sobre algún aspecto o incidencia que superase su capacitación profesional para analizarlo en idóneas condiciones, deberá remitir el mismo a un especialista en la materia o recabar su dictamen para reforzar la calidad y fiabilidad global de la auditoría.*

Dicha definición viene respaldada especialmente por la organización por excelencia que define las líneas maestras en el terreno de la auditoría de SI: ISACA.

ISACA es un proveedor de conocimiento, certificaciones, comunidad, apoyo y educación en seguridad y aseguramiento de sistemas de información, gobierno empresarial, administración de TI así como riesgos y cumplimiento relacionados con TI. Fundada en 1969 desarrolla estándares internacionales de auditoría y control de sistemas de información que ayudan a sus miembros a garantizar la confianza y el valor de los sistemas de información. Asimismo, certifica los avances y habilidades de los conocimientos de TI a través de la mundialmente respetada Certified Information Systems Auditor (Auditor Certificado en Sistemas de Información) (CISA).

ISACA actualiza continuamente COBIT, que ayuda a los profesionales y líderes empresariales de TI a cumplir con sus responsabilidades de administración y gestión, particularmente en las áreas de aseguramiento, seguridad, riesgo y control, para agregar valor al negocio. COBIT enfatiza el cumplimiento regulatorio, ayuda a las organizaciones a incrementar su valor a través de las tecnologías, y permite su alineamiento con los objetivos del negocio.

COBIT ha ido evolucionando en el tiempo, y actualmente se encuentra en su versionado 5, como podemos ver en el siguiente diagrama indicativo de dicha evolución:



**Ilustración 7. Evolución de COBIT (ISACA)**

ISACA desarrolló COBIT 5 para cubrir las necesidades de los interesados, y alinearse a las actuales tendencias sobre técnicas de gobierno y administración relacionadas con la TI. (Cabe recalcar que la práctica de este estándar es mantenido por una organización sin fines de lucro con un gran reconocimiento mundial, y se encuentra totalmente mapeado con otros estándares internacionales como la norma ISO 9001: 2008).

Dicho marco tecnológico se apoya en cinco principios fundamentales, como se puede observar a continuación:

1. Satisfacer las necesidades de las Partes Interesadas.
2. Cubrir la Compañía de Forma Integral.
3. Aplicar un solo Marco Integrado.
4. Habilitar un Enfoque Holístico.
5. Separar el Gobierno de la Administración.



## **5.12 El coste de la no calidad de datos**

“Raramente pueden encontrarse en el mercado estudios sistemáticos del coste de la no calidad de los datos. Es de suponer que esta carencia es debida a que el "problema de los datos" es interpretado como un problema de difícil o imposible solución, tanto por gestores como por técnicos, por lo que se asume una cierta impotencia.

De este modo, al considerarlo un problema irresoluble o de alto coste, se da una cierta "manga ancha" en todos los proyectos que implican manejar los datos de clientes, empleados o productos del catálogo. Se asume la falta de calidad de los datos como algo con lo que hay que convivir, que no tiene solución.

El resultado es claro: se sigue arrastrando una ingente cantidad de datos incorrectos, desactualizados o simplemente inexistentes, produciendo enormes bolsas de ineficiencia en todos los procesos de negocio, desde personal hasta ventas. ¿Es esta una situación aceptable para las empresas?

Según el pasado informe de Gartner de Agosto de 2009 las empresas objeto del estudio de este aspecto, dieron la cifra promedio de 8,2 Millones de Dólares de pérdidas debidas a la no calidad de los datos de clientes. El 20% de los encuestados afirmaban que las pérdidas se aproximaban más a los 20 Millones de Dólares.

Y, viendo estos datos, ¿puedo combatir la no calidad de los datos? ¿Puedo combatirla en solitario?

Combatir la calidad de datos no solo es posible sino que debe ser una prioridad estratégica en toda empresa, pues siempre produce una gran diferenciación respecto de su competencia.

Sin embargo hacerlo en solitario es algo complicado pues las herramientas necesarias son complejas y caras de desarrollar y las que incorporan técnicas de Inteligencia Artificial requieren entrenamiento, cosa que Ud. difícilmente podrá abordar.

Por ello, recomendamos que acuda a empresas verdaderamente especialistas en Calidad de Datos, como Sigma Data Services, quienes podrán diagnosticar el estado de sus datos y podrán ayudarle fácilmente a mejorar la calidad de sus datos dentro de un presupuesto controlado con máximas garantías de resultados.”

(Miguel, 2011, <http://marketing.sigma-data.com/noticias/4-noticias/17-el-coste-de-la-no-calidad-de-datos>)

### 5.13 Artículo de opinión: Las 10 causas más frecuentes de la no calidad de datos

“Son muchas las empresas que nos vienen preguntando desde hace tiempo cuáles son los orígenes de la pérdida de calidad de datos. El interés de esta cuestión radica en que un diagnóstico más preciso de dichas causas puede permitir trazar una política interdepartamental más completa que cuente con el apoyo de la dirección. Por ello, decidimos realizar un estudio sistemático entre algunos de nuestros principales clientes.

Los resultados son realmente sorprendentes, pues históricamente habíamos creído que la mayor parte de problemas de calidad de datos se originaban en las entradas manuales internas, por ejemplo en la grabación de cupones y procesos de este estilo. Aunque muchos problemas tienen esa causa, es interesante observar que la primera causa de problemas críticos proviene de procesos externos, tales como las entradas de datos desde sitios web.”

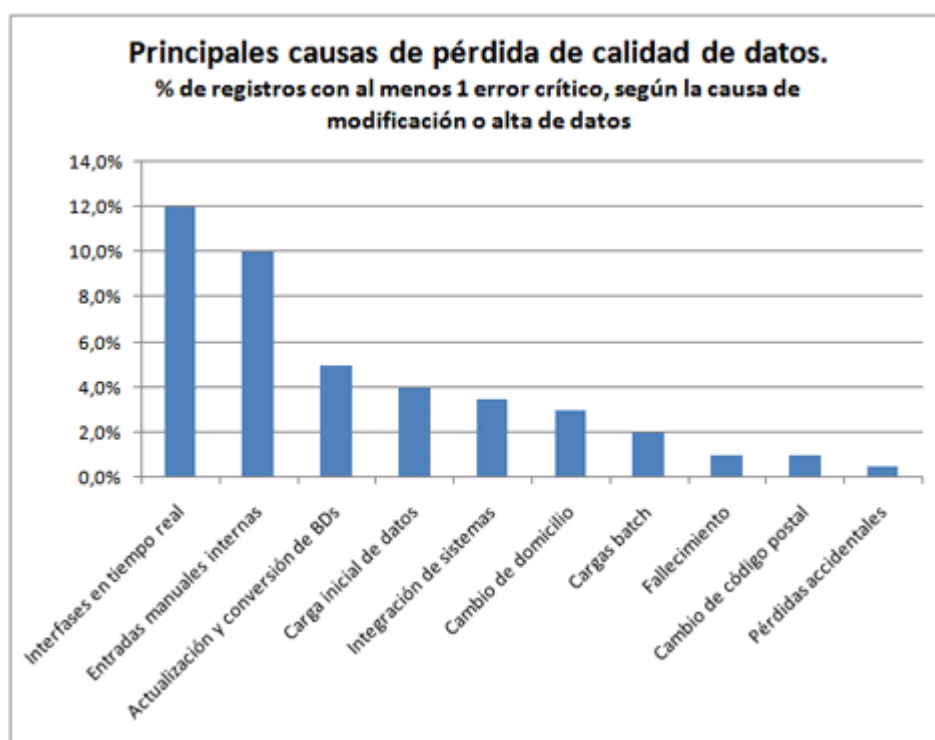


Ilustración 8. Principales causas de pérdida de calidad de datos

(2011, <http://marketing.sigma-data.com/noticias/4-noticias/14-las-10-causas-mas-frecuentes-de-la-no-calidad-de-datos>)

# Implicaciones legales

## ***6.1 ¿Qué es la Ley Orgánica de Protección de Datos de carácter personal?***

La Ley Orgánica 15/1999 de 13 de diciembre de Protección de Datos de Carácter Personal, (LOPD), es una Ley Orgánica española que tiene por objeto garantizar y proteger, en lo que concierne al tratamiento de los datos personales, las libertades públicas y los derechos fundamentales de las personas físicas, y especialmente de su honor, intimidad y privacidad personal y familiar.

Su objetivo principal es regular el tratamiento de los datos y ficheros, de carácter personal, independientemente del soporte en el cual sean tratados, los derechos de los ciudadanos sobre ellos y las obligaciones de aquellos que los crean o tratan.

## ***6.2 Antecedentes normativos***

- La Constitución Española de 1978:

En el artículo 18.4 se dispone:

"La Ley limitará el uso de la informática para garantizar el honor y la intimidad personal y familiar de los ciudadanos y el pleno ejercicio de sus derechos"

- La Ley Orgánica 5/92 de Regulación del Tratamiento Automatizado de los Datos de Carácter Personal de 29 de octubre (LORTAD): No es hasta ese año 1992 cuando nace la primera Ley española que regula de forma específica esta materia.



### Auditoría de calidad de bases de datos de carácter personal

En la actualidad se encuentra derogada, su ámbito de aplicación se circunscribía únicamente a los ficheros de carácter personal que se tratan en soportes automatizados.

- La Directiva Europea 95/46 CE de 24 de octubre del Parlamento Europeo y Consejo relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos.

Finalmente, es a finales de 1999, cuando se publica la Ley Orgánica de Protección de Datos, que amplía el ámbito de aplicación a todo tipo de ficheros, independientemente del soporte en el cual sean tratados. Asimismo se adecua a lo establecido en la Directiva Europea 95/46 CE. Su ámbito de aplicación se circunscribe a todo tipo de ficheros de carácter personal, independientemente del tipo de soporte en el cual se traten (automatizados y no automatizados).

### **6.3 *Ámbito de aplicación.***

La ley Orgánica (15/1999) de Protección de Datos de Carácter Personal (LOPD) afecta a cualquier profesional liberal, empresa u organización pública o privada que guarde algún dato de carácter personal (DCP), como nombres o apellidos, de cualquier persona, sea empleado, cliente, incluso proveedor...Independientemente de si se guardan en ficheros automatizados, como si se guardan en papel. En definitiva, afecta a la totalidad de las empresas.

#### **6.4 ¿Qué dice la Agencia Española de Protección de Datos sobre la calidad de datos?**

La Ley Orgánica 15/1999 contiene entre sus principios generales, el principio de calidad de los datos, que, ligado al principio de proporcionalidad de los datos, exige que los mismos sean adecuados a la finalidad que motiva su recogida.

La recogida y tratamiento de datos de carácter personal debe efectuarse desde su subordinación a los principios de calidad de los datos y de proporcionalidad que establece la Ley.

No puede obviarse que estamos tratando de un auténtico derecho fundamental, cuyo contenido el Tribunal Constitucional ha terminado de perfilar en la Sentencia 292/2000, de 30 de noviembre, denominándolo derecho de autodeterminación informativa o de libre disponibilidad de los datos de carácter personal. Así, en dicha sentencia se indica que este derecho fundamental 'persigue garantizar a esa persona el poder de control sobre sus datos personales, sobre su uso y destino, con el propósito de impedir su tráfico ilícito y lesivo para la dignidad y derecho del afectado', estableciendo, en cuanto a su ámbito, que 'el objeto de protección del derecho fundamental a la protección de datos no se reduce sólo a los datos íntimos de la persona, sino a cualquier tipo de dato personal, sea o no íntimo, cuyo conocimiento o empleo por tercero pueda afectar a sus derechos sean o no fundamentales, porque su objeto no es sólo la intimidad individual, que para ello está la protección que el artículo 18. 1 CE otorga, sino los datos de carácter personal'.

Aún concretando más el contenido del derecho, se establece que el poder de disposición y control sobre los datos personales que tal derecho implica 'se concretan jurídicamente en la facultad de consentir la recogida, la obtención y el acceso a los datos personales, su posterior almacenamiento y tratamiento, así como su uso o usos posibles, por un tercero, sea el Estado o un particular. Y ese derecho a consentir el conocimiento y el tratamiento, informático o no, de los datos personales, requiere como complementos indispensables, por un lado, la facultad de saber en todo momento quién dispone de esos datos personales y a qué uso los está sometiendo, y, por otro lado, el poder oponerse a esa posesión y usos'.

## **6.5 Artículo 4 de la LOPD de la AEPD: calidad de datos**

**Para todas las empresas, que los datos que tratan y almacenan se encuentren actualizados es realmente importante; la no actualización de los datos puede suponer costes económicos importantes ya que disminuye la eficacia en la toma de decisiones, acciones comerciales, nuevos productos o, simplemente, para el seguimiento o mantenimiento de relaciones contractuales.**

*La Ley tiende a establecer unos principios generales de Calidad tendentes a garantizar un uso adecuado de los datos, fijando que:*

- 1º.- Los datos personales deben de adecuarse a la finalidad para la que fueron recabados,*
- 2º.- Deben ser exactos y actualizados,*
- 3º.- No deben mantenerse indefinidamente sin justificación, y*
- 4º.- Deben haber sido recogidos de forma lícita.*

*El incumplimiento de estas obligaciones supone una infracción grave, tipificada en el art.44.3 d), que lleva aparejada sanciones económicas de entre 40.001 a 300.000 €.*

*(“El artículo 45 de la LOPD, en sus apartados 1 a 5, establece, según la nueva redacción dada por la Ley 2/2011, que:*

- 1. Las infracciones leves serán sancionadas con multa de 900 a 40.000 euros.*
- 2. Las infracciones graves serán sancionadas con multa de 40.001 a 300.000 euros.*
- 3. Las infracciones muy graves serán sancionadas con multa de 300.001 a 600.000 euros.”)*

### *a) Finalidad.*

*Significa que los datos sólo se podrán recoger para su tratamiento, así como someterlos a dicho tratamiento, cuando sean adecuados, pertinentes y no excesivos en relación con el ámbito y las finalidades determinadas, explícitas y legítimas para los que se hayan obtenido.*

*Estos criterios de adecuación se relacionan expresamente con las finalidades para las que se han recabado los datos; en el caso que las finalidades del tratamiento cambien, los datos deberán ser cancelados. Por ejemplo, si recogemos datos para la participación en un concurso, dichos datos no los podremos destinar luego a finalidades distintas, diferentes o incompatibles.*

*Por ello, es importante a la hora de recoger los datos determinar en las cláusulas de información todas las finalidades a las que se van a destinar los datos.*





b) *Utilización no abusiva.*

*La utilización no abusiva de los datos impide que los datos se usen para finalidades incompatibles o distintas con aquellas para las que hubiesen sido recogidos, si bien no se considera incompatible el tratamiento posterior de estos datos con fines históricos, estadísticos o científicos.*

c) *Exactitud.*

*Significa que los datos tienen que ser exactos y puestos al día de forma que respondan con veracidad a la situación actual del afectado. Ello no significa que las empresas deban mantener exactos los datos cuando no tengan medios para conocer la exactitud/veracidad de los datos; pero, si tienen conocimiento de la inexactitud de un dato, deben proceder a actualizarlo.*

*Así, la corrección/actualización de los datos tratados puede realizarse:*

*-De oficio por el Responsable del Fichero, cuando conozca dicha inexactitud.*

*-A través del ejercicio del derecho de rectificación de los datos por el propio interesado.*

*En cuanto a datos recabados de fuentes accesibles al público (repertorios telefónicos, boletines oficiales, publicaciones) se mantiene esta obligación de exactitud y veracidad; pero, hay que decir que no es responsabilidad del Responsable del Fichero comprobar si los datos publicados en dichas fuentes son exactos o inexactos, aunque si se conociese la inexactitud debe procederse a actualizarlo.*

d) *Cancelación en caso de no necesidad.*

*El **art.4.5 dispone** que –“Los datos de carácter personal serán cancelados cuando hayan dejado de ser necesarios o pertinentes para la finalidad para la cual hubieren sido recabados o registrados”-.*

*Así, la conexión entre los datos y la finalidad que motivó su recogida se mantiene en todo momento. La cancelación se producirá directamente por el Responsable del Fichero en el momento en que el dato no sea necesario, sin perjuicio de la posibilidad que tiene el afectado de solicitar la cancelación del dato.*

*En el caso de que alguna **obligación legal** establezca la necesidad de conservar los datos una vez concluida la finalidad que motivó su recogida, el Responsable del Fichero podrá conservar los datos a través del boqueo del dato o previo proceso de disociación.*

*En cuanto al bloqueo de los datos, **la LOPD establece**, en el art. 16.3, que –“La cancelación dará lugar al bloqueo de los datos, conservándose únicamente a disposición de las Administraciones Públicas, Jueces y Tribunales, para la atención de las posibles responsabilidades nacidas del tratamiento, durante la prescripción de éstas. Cumplido el citado plazo deberá procederse a su supresión”-.*



## Auditoría de calidad de bases de datos de carácter personal

*El art. 16.5 establece que -“Los datos de carácter personal deberán ser conservados durante los plazos previstos en las disposiciones aplicables o, en su caso, en las relaciones contractuales entre la persona o entidad responsable del tratamiento y el interesado”-.*

*En cuanto al procedimiento de disociación, podemos definirlo como todo tratamiento de datos personales de modo que la información que se obtenga no pueda asociarse a persona identificada o identificable; en la práctica se traduce en la disociación de las tablas y campos que conforman el fichero, de manera que no son relacionables los datos o informaciones obtenidas con una persona identificada o identificable.*

### *e) Almacenamiento.*

*Los datos serán conservados de manera que permitan el ejercicio de los derechos de acceso, salvo que sean legalmente cancelados.*

*El derecho de acceso consiste en la facultad que tiene el afectado de solicitar y obtener, de manera gratuita y en un plazo determinado, la información sobre sus datos sometidos a tratamiento, las comunicaciones que se han realizado y las cesiones realizadas, principalmente. Este derecho será desarrollado con más profundidad en un apartado específico de esta Guía.*

### *f) Lealtad.*

*Se impone la prohibición de recoger los datos por medios fraudulentos, desleales o ilícitos.*

### *g) Incumplimiento de los principios de Calidad de los Datos.*

*La ley establece que la vulneración de los principios y garantías desarrollados en el art.4 tendrán la consideración de infracciones graves o muy graves. Así, encontramos:*

#### **1º.- Infracciones graves:**

**•Art.44.3.b:** *-“Proceder a la creación de ficheros de titularidad privada o iniciar la recogida de datos para los mismos con finalidades distintas de las que constituyen el objeto legítimo de la empresa o entidad”-.*

**•Art.44.3.d:** *-“Tratar los datos de carácter personal o usarlos posteriormente con conculcación de los principios y garantías establecidos en la Ley o con incumplimiento de los preceptos que impongan las disposiciones reglamentarias de desarrollo, cuando no constituyan infracción muy grave”-.*

**•Art.44.3.e:** *-“El impedimento o la obstaculización del ejercicio de los derechos de acceso y oposición y la negativa a facilitar la información que sea solicitada”-.*

**•Art.44.3.f:** *-“Mantener los datos de carácter personal inexactos o no efectuar las rectificaciones o cancelaciones de los mismos que legalmente procedan cuando resulten afectados los derechos de las personas que la presente ley ampara”-.*

#### **2º.- Infracciones muy graves:**



•**Art.44.4.a:** -*“La recogida de datos de forma engañosa y fraudulenta”*

(2011, [http://www.microsoft.com/business/smb/es-es/guia\\_lopd/calidad\\_datos.msp](http://www.microsoft.com/business/smb/es-es/guia_lopd/calidad_datos.msp))

## **6.6 Diversas doctrinas y sentencias reducidas sobre la calidad de datos.**

Para verificar este apartado, a modo de ejemplo, se adjuntan una serie de sentencias reducidas, no extensas, sobre la calidad de los datos de carácter personal. Dichas sentencias están extraídas de la autoridad de la Agencia relativas a la Comunidad de Madrid.

Sentencia de la Sala de lo Contencioso Administrativo de la Audiencia Nacional, de 6 de julio de 2001.

- 'La conducta de la entidad recurrente recabando o intentando recabar unos datos de carácter personal (en concreto los relativos a la cuenta bancario o VISA) para su tratamiento automatizado que resultaban completamente innecesarios e inadecuados en relación con el ámbito y finalidades legítimas para las que se hayan obtenido, debe ser constitutiva de infracción (....) Y aunque los datos no llegaron a ser incorporados a los ficheros (...), ello no implica que falte el necesario tratamiento automatizado de los mismos para que se produzca el tipo sancionador.'

Sentencia de la Sección Novena de la Sala de lo Contencioso Administrativo del Tribunal Superior de Justicia de Madrid, de 5 de noviembre de 1998.

- 'Sin embargo, aunque (...) no hubo (...) una intención de dañar ni enriquecimiento injusto (...) los hechos han tenido una doble perturbación para la perjudicada: (...) imputarle una deuda inexistente (.....) lo más grave fue su inclusión en un Registro Informático de Morosos y además sin conocimiento de la perjudicada (....) y de esa inclusión indebida en el Registro de Morosos no eran responsables los que llevan el Registro sino los que suministraron el dato'.

Sentencia de la Sección Octava de la Sala de lo Contencioso Administrativo del Tribunal Superior de Justicia de Madrid, de 18 de octubre de 2000.

- '(....) para incluir en un fichero de solvencia patrimonial el dato relativo a una deuda, ésta, además de cierta, vencida y exigible, ha de haber resultado efectivamente impagada (....) debiendo además, el acreedor (como requisito previo a la inclusión del dato en un fichero de estas características) proceder en la forma más arriba descrita y cuya finalidad no es otra que garantizar la exactitud de los datos que se pretender incluir'.

Sentencia de la Sección Octava de la Sala de lo Contencioso Administrativo del Tribunal Superior de Justicia de Madrid, de 26 de mayo de 1999.



### Auditoría de calidad de bases de datos de carácter personal

- 'Si los datos registrados se han obtenido de una fuente accesible al público (como en el caso de autos) el medio más efectivo para mantener actualizados aquellos será notificando al afectado la existencia del dato a fin de que éste (si el dato obtenido de esa fuente de acceso público es incorrecto o la situación ha variado) pueda instar las rectificaciones pertinentes en el momento en que el dato registrado no responda a la realidad y si el afectado declina realizar las oportunas rectificaciones, entendemos, su inactividad exculpará al titular del fichero de toda responsabilidad en orden a la actualización de los datos, en la medida que esa actualización no pueda obtenerse de la misma forma en la que se obtuvo el dato'.

Sentencia de la Sala de lo Contencioso Administrativo de la Audiencia Nacional, de 19 de enero de 2001.

- (...) incluido el dato erróneo, la infracción se produce hasta que el mismo haya sido erradicado del fichero'.

Sentencia de la Sección Octava de la Sala de lo Contencioso Administrativo del Tribunal Superior de Justicia de Madrid, de 9 de febrero de 2000.

- '(...) el dato registrado (...) es transcripción del contenido en dos edictos publicados en el BOCAM, por lo que ignoramos si son o no exactos dichos datos y, en todo caso, la inexactitud del mismo nunca sería imputable a la actora. Si (...) para los datos obtenidos de fuente accesibles al público la LORTAD no exige la notificación del registro al afectado, difícilmente puede saber el titular del fichero si el dato obtenido de una fuente accesible al público es o no correcto y, además, en el caso de autos, dado que en el edicto no consta otro datos que el nombre y apellidos de los demandados, nunca hubiera sido posible efectuar tal notificación, ni averiguar la exactitud del dato publicado, ni de lo actuado puede afirmarse que dicho dato se refiera siquiera al denunciante, por lo que en la medida que no conste al titular del fichero la inexactitud del dato registrado, inexactitud que, reiteramos, no consta, no existe para éste la obligación legal de cancelar el dato'.

Sentencia de la Sala de lo Contencioso Administrativo de la Audiencia Nacional, de 9 de marzo de 2001.

- 'Uno de los principios que inspira la legislación sobre tratamiento automatizado de datos de carácter personal es el de calidad de datos. Este principio implica, entre otras cosas, que los datos sean necesarios y pertinentes para la finalidad para la cual hubieran sido recabados o registrados (art. 4.5 de la LO 5/1992) y que sean exactos y completos art. 4.4 de la LO 5/1992. Por lo tanto, si los datos han dejado de ser necesarios para los fines para los cuales fueron recabados o registrados o resultan inexactos, se debe proceder (...) a su cancelación, sin necesidad de solicitud del afectado. Y así se infiere del propio tenor literal de los artículos 4.4 y 4.5 de la LO 5/1992, que utiliza la expresión imperativa 'serán cancelados' y sin condicionarla a la existencia de una previa solicitud del afectado. En suma, la norma establece la obligación del responsable del fichero de proceder de oficio y con la debida diligencia a cancelar los datos inexactos o que han dejado de ser necesarios para la finalidad del fichero y sin necesidad de solicitud previa del afectado'.

## 6.7 La importancia de la recogida de datos de carácter personal

Los datos recogidos serán adecuados, pertinentes y no excesivos atendiendo a la finalidad para la que han sido recabados.

Los datos de carácter personal:

- No podrán usarse para finalidades incompatibles para las que fueron recogidos.
- Serán exactos y puestos al día.
- Serán cancelados cuando dejen de ser necesarios y cuando no sean exactos o estén incompletos.
- Salvo que se mantengan para valores históricos, científicos estadísticos, una vez que dejen de ser útiles para la función requerida no podrán ser conservados.
- No se podrán recoger los datos por medios fraudulentos, desleales o ilícitos.

## 6.8 Implicaciones legales a nivel europeo

Con relación a las implicaciones legales en el marco europeo, el borrador del Reglamento del Parlamento Europeo y del Consejo, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos (Reglamento general de protección de datos), del año 2012, posee en su artículo 43 el siguiente articulado:

*Transferencias mediante normas corporativas vinculantes.*

*Las normas corporativas vinculantes especificarán, como mínimo:*

- a) la estructura y los datos de contacto del grupo de empresas y de sus miembros;*
- b) las transferencias o serie de transferencias de datos, incluidas las categorías de datos personales, el tipo de tratamientos y sus fines, el tipo de interesados afectados y el nombre del tercer o los terceros países en cuestión;*
- c) su carácter jurídicamente vinculante, tanto a nivel interno como externo;*
- d) los principios generales en materia de protección de datos, en particular la limitación de la finalidad, la **calidad de los datos**, la base jurídica del tratamiento, el tratamiento de datos personales sensibles, las medidas encaminadas a garantizar la seguridad de los datos y los requisitos en materia de transferencias ulteriores a organizaciones que no estén vinculadas por esas políticas;*



# Soluciones comerciales

En este capítulo expondremos algunas de las empresas actuales que tienen soluciones de auditoría de calidad en bases de datos, así como algoritmos de normalización de datos.

En la actualidad, existen diversas empresas cuyo trabajo es sanear las bases de datos existentes de sus clientes, y, hasta en algunas, figuran lemas como “La calidad de sus datos es nuestra Responsabilidad”, pero ahora bien, por lo visto hasta ahora, existen muy pocas empresas que se encarguen del estudio previo de los datos que se van a introducir en éstas bases de datos.

La solución aportada es una herramienta de ayuda al auditor en su labor, para medir el nivel de salud, o calidad, con que se van a introducir una serie de datos en una base de datos.

A continuación se hace una pequeña referencia a cinco empresas del sector de la calidad de los datos, en el que veremos brevemente la función que desempeñan:

### 7.1 DEYDE

En DEYDE se dedican a servicios de normalización, deduplicación y enriquecimiento para que las empresas puedan comunicarse eficazmente con sus clientes.

Su labor principal es la de normalización de nombre y direcciones además de la deduplicación de datos.



Ilustración 9. Soluciones DEYDE

En su sitio web, incluyen una pequeña demostración:





**Ilustración 10. Soluciones DEYDE 2.**

Aunque, esta solución de datos que ofrecen está diseñada para extraer los datos de la base de datos de sus clientes, y corregir o normalizar el formato de éstos. No ofrece una solución para el paso anterior de normalizar los datos antes de introducirse en sus bases de datos respectivas, tema principal del aporte de este Proyecto.

(2012, [www.deyde.es](http://www.deyde.es))



## 7.2 CAMERDATA

Esta empresa se encarga de la actualización de las bases de datos de sus clientes, proporcionan las altas, bajas y modificaciones de los ficheros.

También se dedican a incorporar nuevos campos en las bases de datos.

Aunque, su principal objetivo es devolver las bases de datos actualizadas con información normalizada y deduplicada, con todos los campos validados.



Ilustración 11. Soluciones CAMERDATA

Como en el caso de la anterior empresa citada (Deyde), Camerdata tiene como objetivo el objetivo de devolver una base de datos actualizada con información normalizada y deduplicada, que puede incorporar nuevos campos y con los datos validados e investigados. En ningún momento hace alusión a normalizar dichos datos antes de introducirlos en las bases de datos de las empresas.

(2012, [www.camerdata.es](http://www.camerdata.es))

### 7.3 SCHOBER

Proporciona soluciones para la mejora de la calidad de las bases de datos de sus clientes.

Dicha empresa ofrece servicios de estandarización, validación, normalización y deduplicación de direcciones para mejorar la calidad de los datos de las bases de datos de sus clientes.



**Ilustración 12. Soluciones SCHOBER**

Schober, comparte la idea de que los datos y la información sobre clientes y mercados son uno de los activos más valiosos con los que cuentan las empresas para generar conocimiento y tomar decisiones estratégicas.

Proporcionan soluciones avanzadas para el tratamiento, depuración y mejora de las bases de datos de clientes, desarrollando servicios de análisis, diseño e implementación de soluciones integradas para la mejora de la calidad de la información en cada proceso de negocio (además incluye la realización de un test de calidad de los datos de las bases de datos de los clientes sin coste alguno y sin compromiso).

Gestionan la información con el objetivo de optimizar el proceso de gestión, *asegurar la calidad de su entrada en las aplicaciones*, analizarla, integrarla y hacer eficiente el proceso de toma de decisiones basada en información precisa, actualizada y completa.

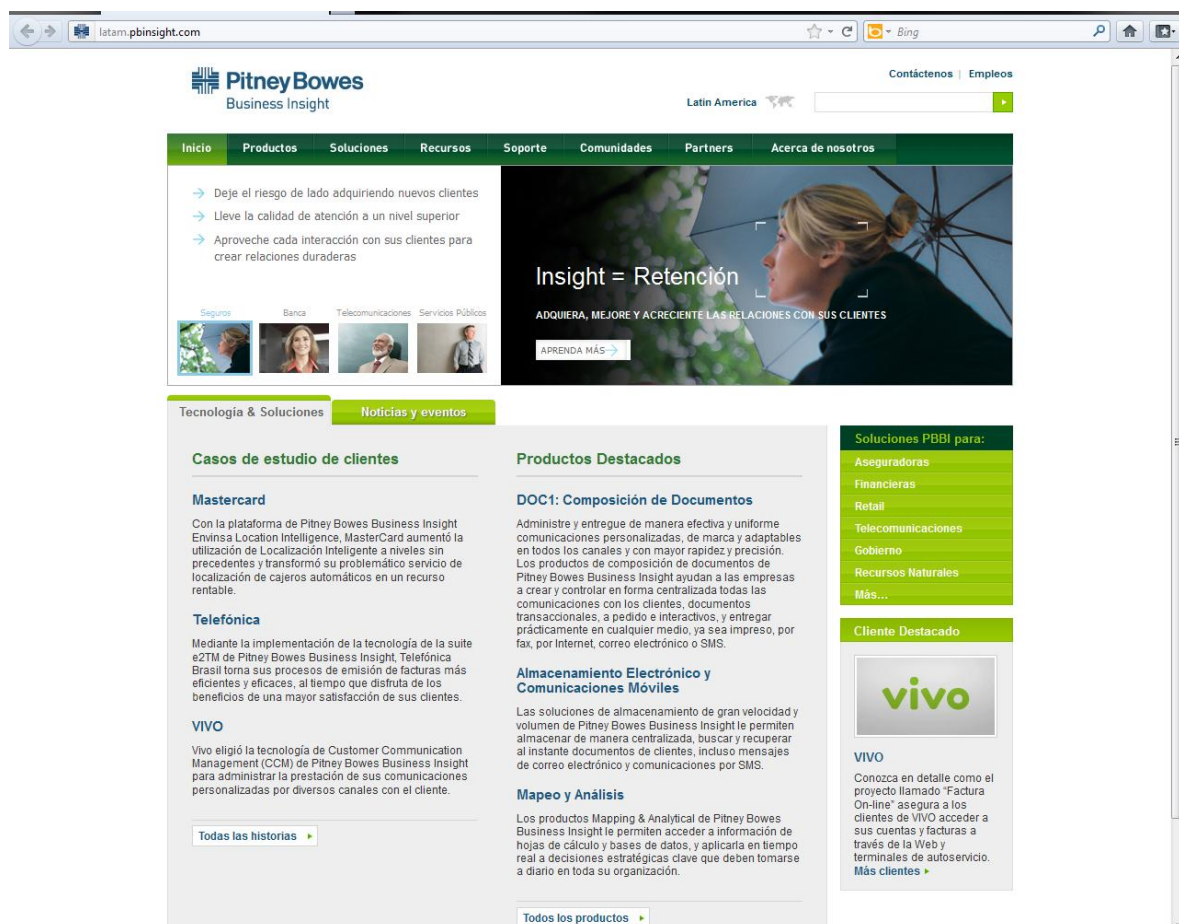
Es decir, esta empresa es una de las pocas del sector que tiene como uno de sus objetivos gestionar los datos en la entrada a las aplicaciones, tema principal de este Proyecto.

(2012, [www.schober.es](http://www.schober.es))

## 7.4 PITNEY BOWES

Esta empresa ofrece soluciones para aumentar el valor de las relaciones con los clientes con una base sólida de calidad de datos. Ofrece ayuda para optimizar los datos de las bases de datos de sus clientes.

Dicha empresa, hace mucha alusión a los costes: “el coste de la escasez de datos puede ser considerable, incluyendo la duplicación de esfuerzos, clientes insatisfechos y pérdida de ingresos”.



**Ilustración 13. Soluciones PITNEY BOWES**

Pitney Bowes proporciona herramientas para crear y mantener actualizada la información consistente. Propone optimizar los procesos empresariales que se basan en datos de buena calidad para obtener información más precisa y fiable.

Esta empresa propone la calidad de datos para localizar nuevas oportunidades, conectar con los clientes y comunicar de maneras más eficiente, aunque tampoco hace referencia a proporcionar herramientas que midan la calidad de los datos antes de introducirlos en las bases de datos de las empresas. (2012, [www.pbinsight.com](http://www.pbinsight.com))

## 7.5 AXESOR

Esta empresa se dedica a asesorar a las empresas clientes sobre la normalización de sus bases de datos, ofrece un servicio de marketing.

Proporciona un servicio de asesoramiento para la normalización de nombres en las bases de datos.

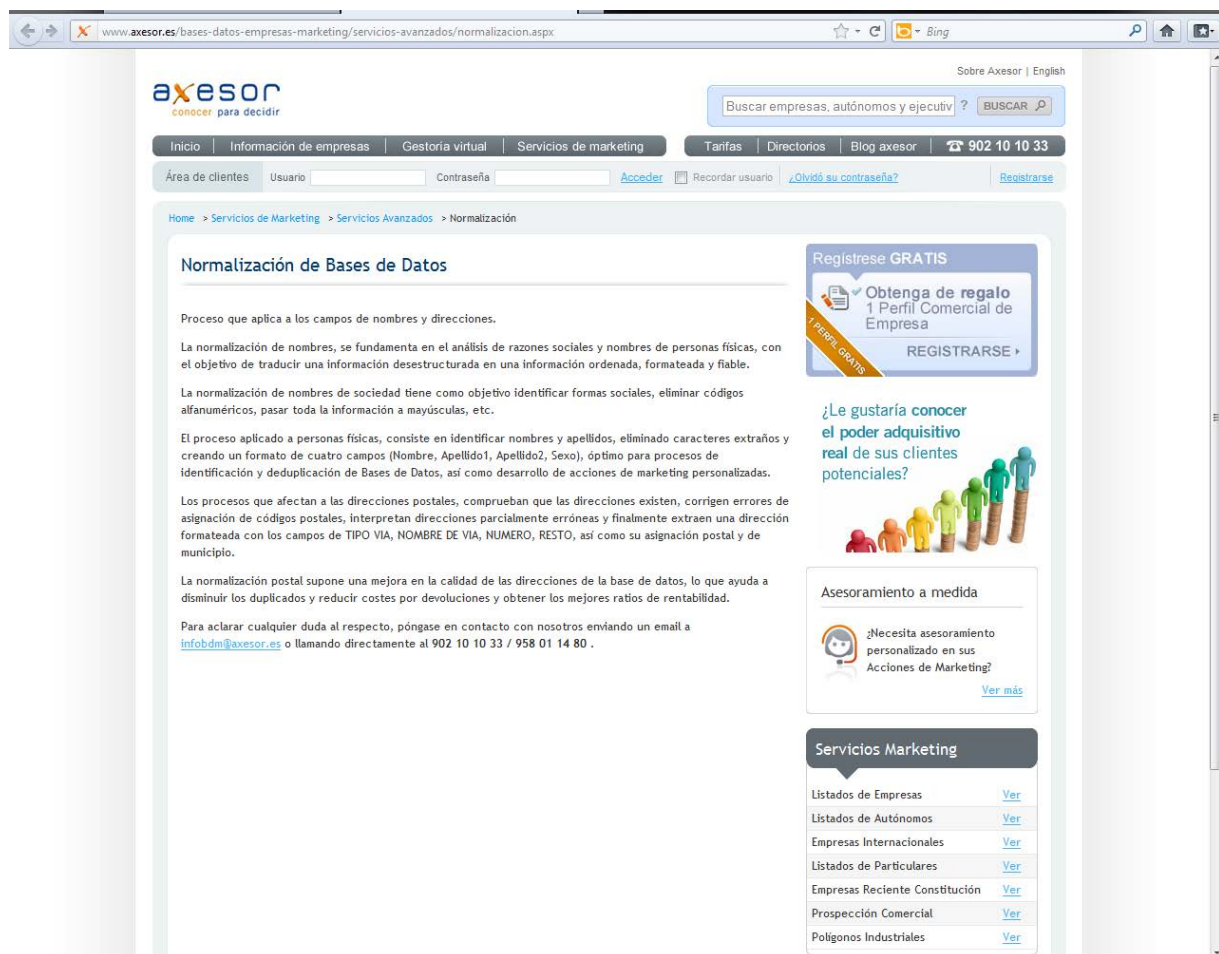


Ilustración 14. Soluciones AXESOR

Esta empresa hace referencia a un estudio que hizo con anterioridad sobre la calidad de los datos:

*“Entre mayo y junio del 2006 se llevó a cabo un estudio con el objetivo de analizar la importancia de los datos y la información como factor clave en la estrategia de las empresas. Se realizaron 1000 entrevistas a ejecutivos con responsabilidades en diversas áreas. Los resultados no dejan de ser sorprendentes.*



### Auditoría de calidad de bases de datos de carácter personal

*La mayoría de los entrevistados piensa que la calidad de los datos es muy importante para su actividad y que la falta de calidad de los mismos afecta de manera importante a la rentabilidad de las acciones. El problema existe y está identificado, es un primer paso. Gran parte de los entrevistados consideraban que la calidad de sus datos era mejorable, pero la sorpresa llega cuando casi el 75% reconocía que la calidad de los datos era peor de lo que pensaban”.*

Aunque a parte de proporcionar dicho servicio de asesoramiento de normalización, no habla del tratamiento de los datos antes de introducirlos en las bases de datos de las empresas.

(2012, [www.axesor.es](http://www.axesor.es))



## 7.6 Soluciones comerciales en el marco internacional

A continuación se muestran algunas de las soluciones internacionales relativas a la calidad de los datos.

### 7.6.1 TALEND\*

Esta empresa estadounidense se dedica a la calidad de la información en general, transmitiendo a las organizaciones la importancia de la misma y mejorando la integridad, exactitud e integridad de sus datos. Para ello, posee dos líneas de trabajo “Talend Open Studio for Data Quality” y “Talend Platform for Data Management”.

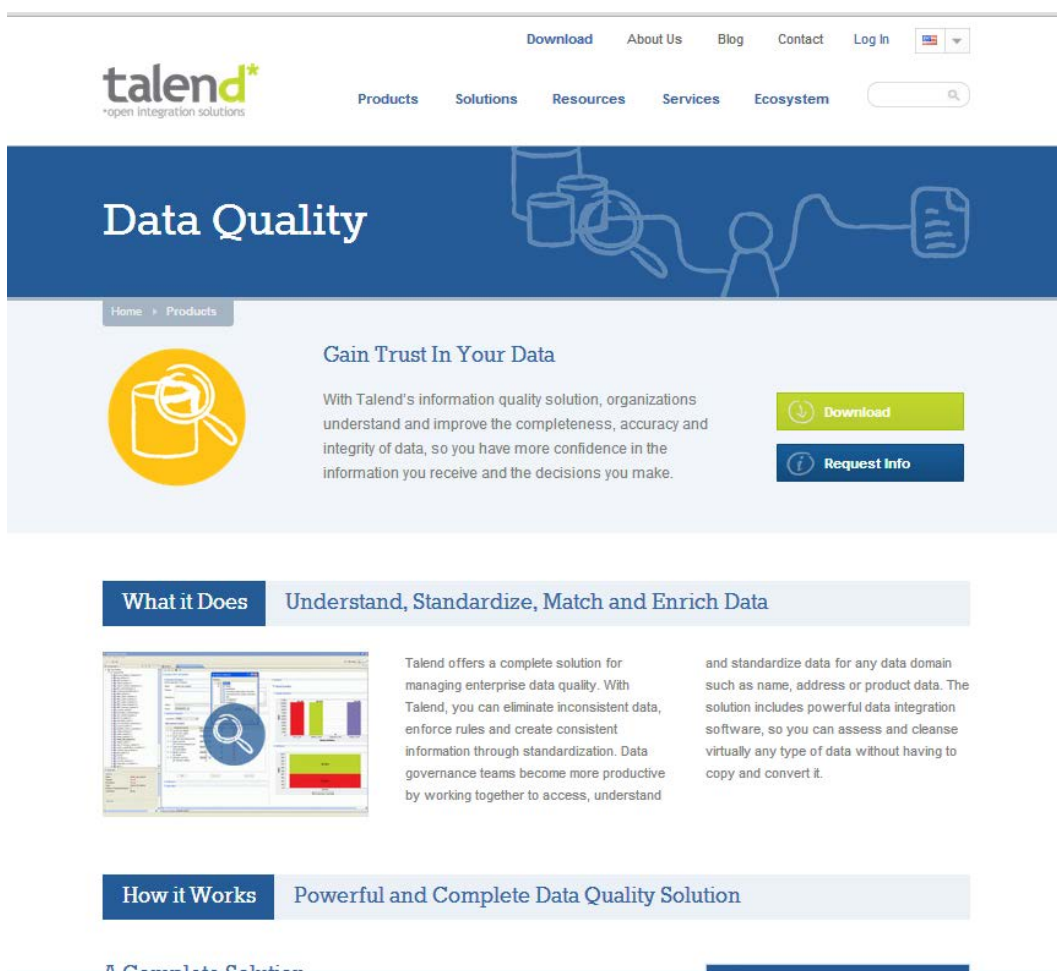
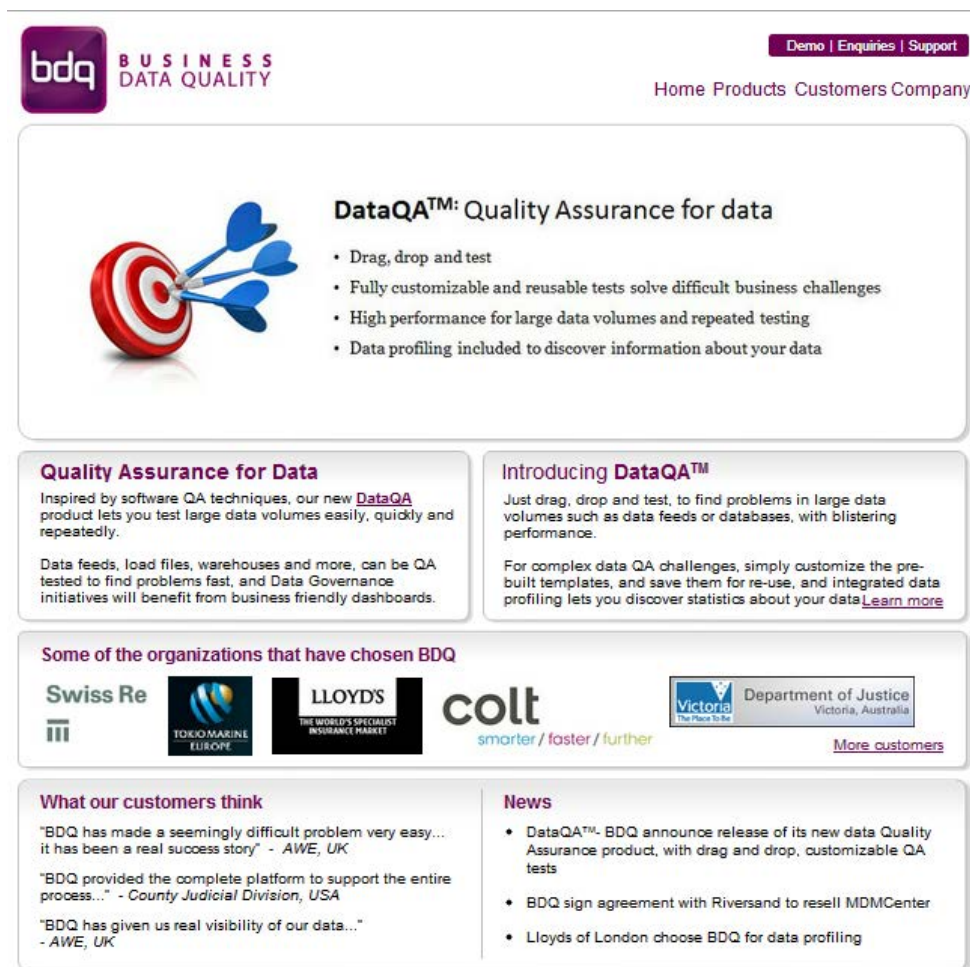


Ilustración 15. Soluciones Internacional Talend\*

(2012, <http://www.talend.com/products/data-quality>)

## 7.6.2 BUSINESS DATA QUALITY

Esta empresa londinense se dedica a la normalización de datos, inspirada por las técnicas de control de calidad de software. Poseen una serie de productos que permiten probar grandes volúmenes de datos de forma rápida, para de este modo, encontrar problemas en las bases de datos.



The screenshot shows the Business Data Quality (BDQ) website. At the top, there is a navigation bar with the BDQ logo, the text "BUSINESS DATA QUALITY", and links for "Demo | Enquiries | Support". Below the navigation bar, there is a main section titled "DataQA™: Quality Assurance for data" featuring an image of a target with three darts hitting the bullseye. To the right of the image is a list of bullet points: "Drag, drop and test", "Fully customizable and reusable tests solve difficult business challenges", "High performance for large data volumes and repeated testing", and "Data profiling included to discover information about your data". Below this, there are two columns of text. The left column is titled "Quality Assurance for Data" and describes the product's inspiration from software QA techniques. The right column is titled "Introducing DataQA™" and describes the product's capabilities for testing large data volumes. Below these columns, there is a section titled "Some of the organizations that have chosen BDQ" featuring logos for Swiss Re, Tokio Marine Europe, Lloyd's, colt, and the Department of Justice Victoria. At the bottom, there are two columns of text. The left column is titled "What our customers think" and contains three quotes from customers. The right column is titled "News" and contains three bullet points about recent announcements.

**bdq BUSINESS DATA QUALITY** Demo | Enquiries | Support

Home Products Customers Company

**DataQA™: Quality Assurance for data**

- Drag, drop and test
- Fully customizable and reusable tests solve difficult business challenges
- High performance for large data volumes and repeated testing
- Data profiling included to discover information about your data

**Quality Assurance for Data**  
Inspired by software QA techniques, our new **DataQA** product lets you test large data volumes easily, quickly and repeatedly.  
Data feeds, load files, warehouses and more, can be QA tested to find problems fast, and Data Governance initiatives will benefit from business friendly dashboards.

**Introducing DataQA™**  
Just drag, drop and test, to find problems in large data volumes such as data feeds or databases, with blistering performance.  
For complex data QA challenges, simply customize the pre-built templates, and save them for re-use, and integrated data profiling lets you discover statistics about your data [Learn more](#)

**Some of the organizations that have chosen BDQ**

Swiss Re  
Tokio Marine Europe  
LLOYD'S  
colt  
Victoria  
Department of Justice  
Victoria, Australia

[More customers](#)

**What our customers think**

"BDQ has made a seemingly difficult problem very easy... it has been a real success story" - AWE, UK  
"BDQ provided the complete platform to support the entire process..." - County Judicial Division, USA  
"BDQ has given us real visibility of our data..." - AWE, UK

**News**

- DataQA™- BDQ announce release of its new data Quality Assurance product, with drag and drop, customizable QA tests
- BDQ sign agreement with Riversand to resell MDMCenter
- Lloyds of London choose BDQ for data profiling

© 2012 Business Data Quality Ltd. | [www.businessdataquality.com](http://www.businessdataquality.com) | Enquiries

Ilustración 16. Soluciones Internacional Business Data Quality

(2012, <http://www.businessdataquality.com/>)

### 7.6.3 DATA QUALITY ONLINE

En este ejemplo, se puede observar como el gobierno australiano, con ayuda del servicio australiano de estadística permiten a los usuarios de esta web a medir la calidad que poseen los datos de sus bases de datos. Tiene desarrollado este servicio para ayudar al ciudadano en su toma de decisiones en general gracias a estos recursos en línea. De este modo, proporcionan información general sobre la calidad de los datos y permiten una posible reutilización de la información por la comunidad de usuarios que lo deseen.

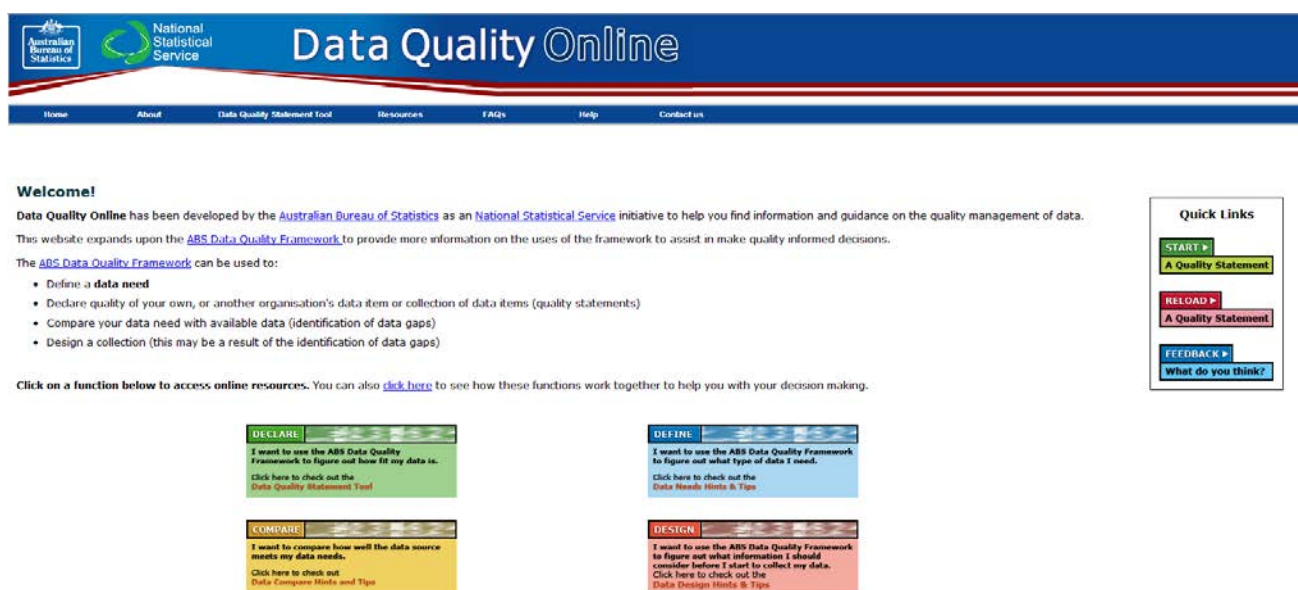


Ilustración 17. Soluciones Internacional Data Quality Online

(2012, <http://www.nss.gov.au/dataquality/>)



## 7.6.4 DATAFLUX

Esta empresa estadounidense se dedica también a la calidad de datos en general, permite la agilidad del negocio y la eficiencia de TI, proporcionando tecnología innovadora para la gestión de datos y servicios que transforman los datos en un activo estratégico. Con esta optimización en la calidad de los datos permite ayudar a reducir los costes, optimizar los ingresos y mitigar los riesgos.

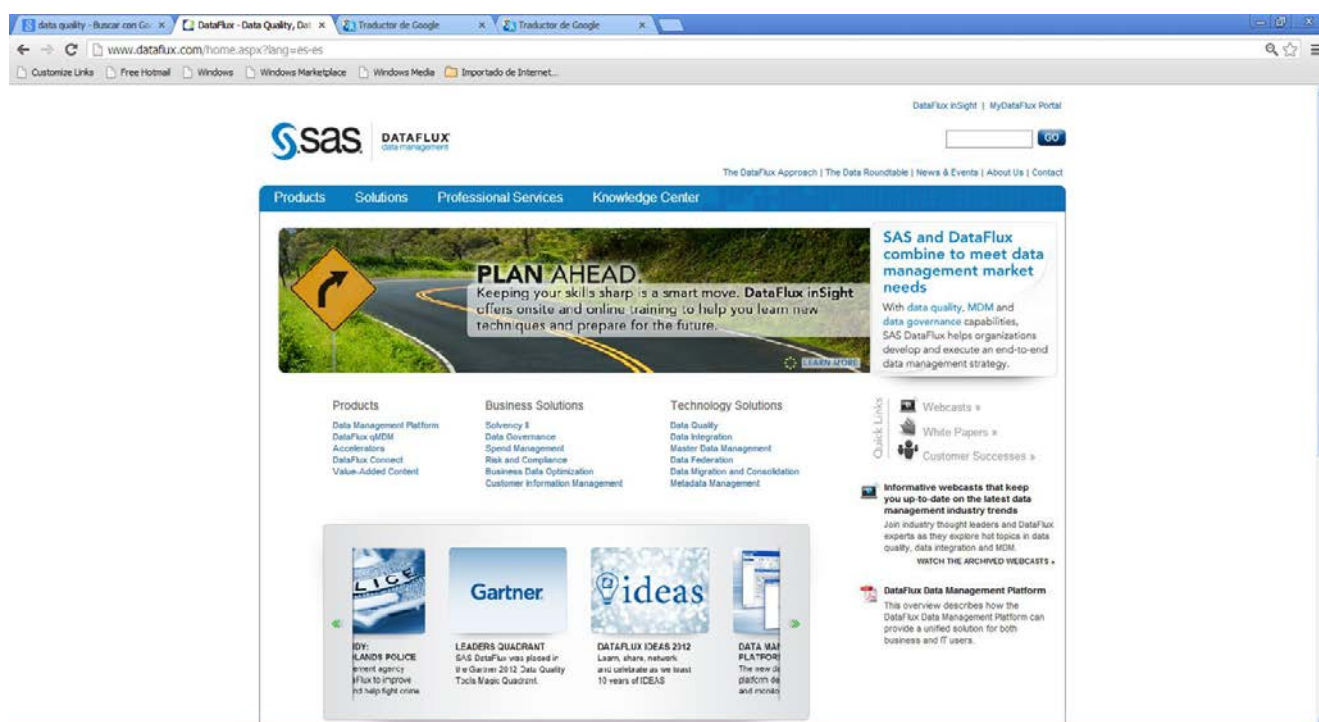


Ilustración 18. Soluciones Internacional Dataflux

(2012, <http://www.dataflux.com/home.aspx?lang=es-es>)



# Bloque III

## Métrica

### 8.1 Introducción.

Una vez mostrado el estado del arte en cuando a la calidad de los datos en la actualidad, e introducido al usuario en el marco teórico y legal del tema tratado, se pretende establecer una métrica que nos permita evaluar el nivel de calidad, o nivel de salud de nuestros datos a introducir en diferentes bases de datos.

Para empezar a mostrar la métrica elegida, definiremos el término métrica del software:

*“Una métrica es cualquier medida o conjunto de medidas destinadas a conocer o estimar el tamaño u otra característica de un software o un sistema de información, generalmente para realizar comparativas o para la planificación de proyectos de desarrollo.”*

Nuestra auditoría de calidad será sobre datos de carácter personal. Una posible definición de estos tipos de datos podría ser la siguiente:

*“Los datos de carácter personal están definidos legalmente como cualquier información numérica, alfabética, gráfica, fotográfica, acústica o de cualquier otro tipo concerniente a personas físicas identificadas o identificables”*

Así pues, los campos de nuestra base de datos son los siguientes (separados si son atributos de identificación o de dirección de una persona):

- Campos de identificación:

- DNI
- Nombre
- Apellido\_1
- Apellido\_2
- Sexo
- Día
- Mes
- Año



- Campos de dirección:

- Calle
- Población
- Provincia
- Código Postal
- Email
- Teléfono Móvil
- Teléfono Fijo

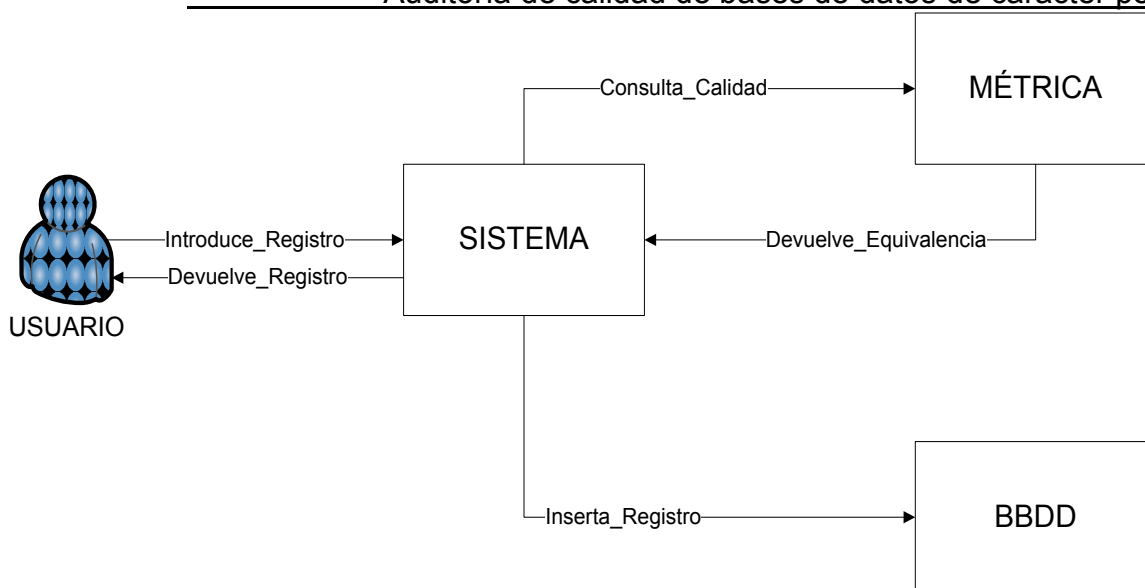
Estos campos estarán disponibles para insertarlos con cuatro diferentes funcionalidades, es decir, el usuario podrá introducir sus registros en cuatro diferentes bases de datos:

- Universidad: contiene los datos de los estudiantes.
- Banco: contiene los datos de los clientes.
- Web de empleo: contiene los datos de los usuarios.
- Videoclub: contiene los datos de los clientes.

Por lo cual, la importancia de nuestros catorce campos variará en función de que el usuario desee introducir los datos en las diferentes bases de datos, así como si dichos campos o atributos son de identificación o de dirección.

El procedimiento del funcionamiento del servicio será el siguiente:

- El usuario introducirá un registro en el sistema, concretamente en un formulario web que contendrá los campos anteriormente citados, y, éste, le pasará al servicio web dicho registro, pues será quien se encargue de consultar con la métrica para saber si se puede introducir en la bbdd dicho registro, siguiendo la tabla de equivalencias. El siguiente gráfico a modo de esquema resume el funcionamiento del servicio:



**Ilustración 19. Esquema gráfico del protocolo de funcionamiento del servicio**

La métrica permitirá al sistema, como hemos dicho, insertar o no el registro en la bbdd si está por encima del umbral que definiremos a continuación. Se ponderarán cada uno de los campos de manera particular según sean para las diferentes bbdd. Una vez hechas estas ponderaciones, el sistema realizará una media (no aritmética) de los datos de identificación y de dirección. Esta medida variará para cada una de las cuatro bases de datos.

Los resultados obtenidos de las cuatro indizaciones expresarán el nivel de calidad o de salud de los datos a introducir en nuestra base de datos y los expondremos dentro de las siguientes equivalencias:

- Óptima:  $x E (91,100)$
- Muy Buena:  $x E (75,90)$
- Buena:  $x E (55,74)$
- Regular:  $x E (26,54)$
- -----u m b r a l-----
- Insuficiente:  $x E (11,25)$
- Pésima:  $x E (0,10)$

(Donde  $x E N$ )

*(Cabe recalcar que todas estas ponderaciones y las que se muestran en toda la métrica podrán ser modificadas a petición del propietario del servicio si así lo requiere)*



## Auditoría de calidad de bases de datos de carácter personal

Es decir, nuestra métrica, expresará el nivel de calidad de los datos en cuanto a su inserción en el aspecto de la identificación, en el de dirección, y la media ponderada de las dos. Por lo cual el resultado de nuestra métrica serán tres valores con los que podremos juzgar el nivel de calidad o la salud que poseen los datos que deseamos insertar en cada una de las cuatro bases de datos expuestas.

Como podemos observar en el pequeño esquema de las equivalencias, para intentar que los datos sean lo más saneados posibles, a la hora de insertarlos en una bbdd, los datos que estén por debajo de la equivalencia “regular”, se devolverán al usuario y no se permitirá insertarlos en dicha bbdd. Es decir el sistema comprobará si la métrica devuelve registros en la escala de equivalencia “pésima” o “insuficiente” y no permitirá su inserción. Se puede decir que el *umbral* a partir del cual no se puede aceptar el registro en la bbdd es el que adjunta a estas dos equivalencias mencionadas.

### 8.2 Grafismo utilizado.

Para medir los valores obtenidos del análisis de los campos de nuestra base de datos, se pueden utilizar muchas clases de grafismo, aunque en este proyecto se utilizarán los diagramas de Kiviat:

*“Un gráfico Kiviat o gráfico radial es un método gráfico para mostrar datos en múltiples variables en la forma de un gráfico de dos dimensiones de tres o más variables representadas en los ejes con el mismo origen. La posición relativa y el ángulo de los ejes son normalmente poco importantes.*

*El esquema también se conoce como web gráfico de araña, mapa de las estrellas, gráfico*

*gráfico consiste en una secuencia de los rayos originan en un centro y formando ángulos iguales cada uno representa una serie de variables. La desde el punto central marcado en el haz es*

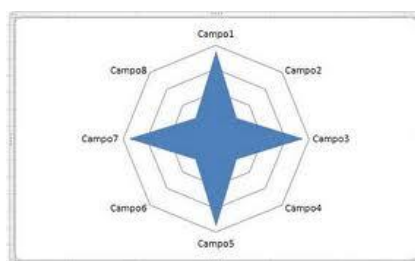


Ilustración 20. Gráfico de Kiviat

*gráficos, polar...*

*Este que se entre sí y distancia*

*proporcional al valor de la*

*variable el valor máximo alcanzado. Los puntos en los radios se unen por segmentos, de modo que la gráfica tiene la forma de una estrella o una tela de araña.*

*La carta de la estrella se utiliza a menudo a la inmediatez con la que podemos comparar las tuplas de valores relacionados con diferentes observaciones.*

*El diagrama Kiviat se utilizó por primera vez en 1877 por Georg von Mayr.”*

### 8.3 Métrica a aplicar para las diferentes indizaciones.

Usando la métrica elegida, llevada a los *diagramas de Kiviat*, se medirán y mostrarán cuatro niveles de calidad diferentes para nuestros datos a introducir. Estos diferentes niveles o indizaciones se utilizarán a medida que el usuario desee introducir campos para las diferentes finalidades (bbdd).

Como se ha dicho en la introducción anterior, la métrica permitirá al sistema insertar o no el registro en la bdd si está por encima del umbral definido. Se ponderarán cada uno de los campos de manera particular según sean para las diferentes bdd, y, una vez hechas estas ponderaciones, el sistema realizará una media, no aritmética, de los datos de identificación y de dirección, como podemos observar a continuación:

- Universidad: Identificación 60%, dirección 40%.
  - En este caso ponderará de esta manera pues para una universidad es más importante los datos de identificación de la persona que los datos referentes a su dirección, para, por ejemplo, subvencionar al alumno.
- Banco: Identificación 70%, dirección 30%.
  - En este caso ponderará de esta manera pues para un banco es mucho más importante los datos de identificación de la persona que los datos referentes a su dirección, para poder tener a la persona totalmente identificada, en caso de solicitar, por ejemplo, un préstamo.
- Web de empleo: Identificación 65%, dirección 35%.
  - En este caso ponderará de esta manera pues para una web de empleo es más importante los datos de identificación de la persona que los datos de su dirección, para poder filtrar, por ejemplo, por titulación obtenida, con total constancia de que se está realizando con total veracidad.
- Videoclub: Identificación 45%, dirección 55%.
  - En este caso ponderará de esta manera pues para un videoclub es menos importante los datos de identificación de la persona que los datos de su dirección, para, por ejemplo, posibles devoluciones.

La importancia de nuestros campos también variará en función de que el usuario desee introducir los datos en las diferentes bases de datos, así como si dichos campos son de identificación o de dirección:

- Si el usuario desea introducir los datos para una base de datos de una universidad, los diferentes campos, ponderados en su totalidad en un 100%, tendrán la siguiente valoración, o peso:



Datos de identificación:

DNI	15
NOMBRE	16
APELLIDO1	17
APELLIDO2	17
SEXO	8
DIA	9
MES	9
AÑO	9

Ilustración 21. Campos de identificación de Universidad

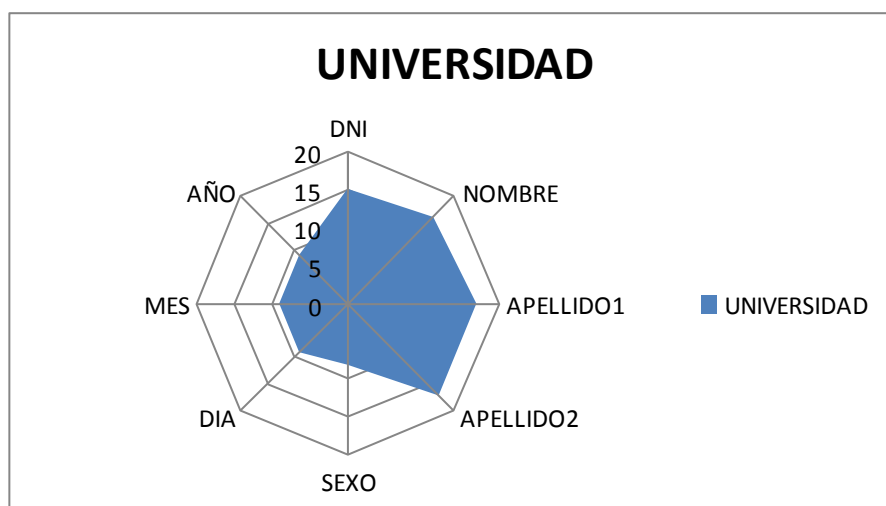
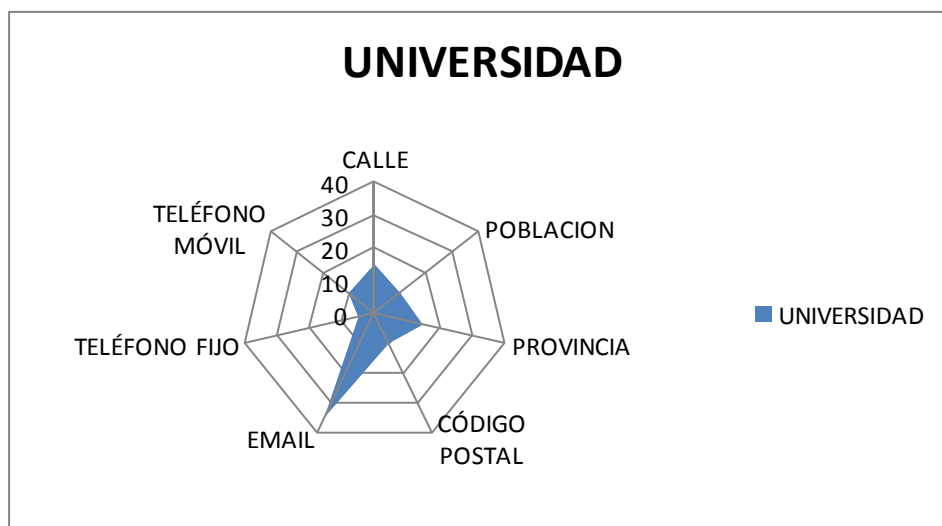


Ilustración 22. Gráfico de identificación de Universidad

Datos de dirección:

CALLE	15
POBLACION	10
PROVINCIA	15
CÓDIGO POSTAL	10
EMAIL	35
TELÉFONO FIJO	5
TELÉFONO MÓVIL	10

Ilustración 23. Campos de dirección de universidad



**Ilustración 24. Gráfico de dirección de Universidad**

- Si el usuario desea introducir los campos para su uso en una base de datos de un banco, la importancia de éstos será:

Datos de identificación:

DNI	21
NOMBRE	17
APELLIDO1	18
APELLIDO2	18
SEXO	5
DIA	7
MES	7
AÑO	7

**Ilustración 25. Campos de identificación de Banco**



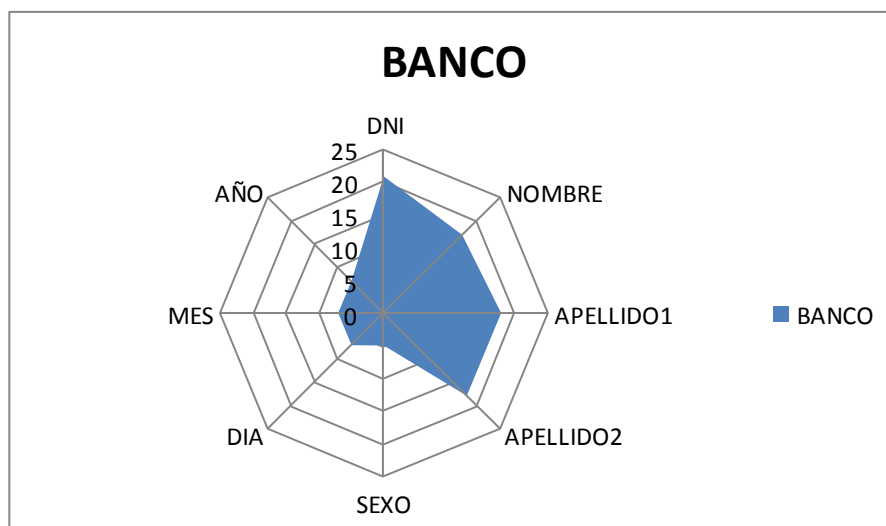


Ilustración 26. Gráfico de identificación de Banco

Datos de dirección:

CALLE	19
POBLACION	16
PROVINCIA	16
CÓDIGO POSTAL	14
EMAIL	6
TELÉFONO FIJO	15
TELÉFONO MÓVIL	14

Ilustración 27. Campos de dirección de Banco

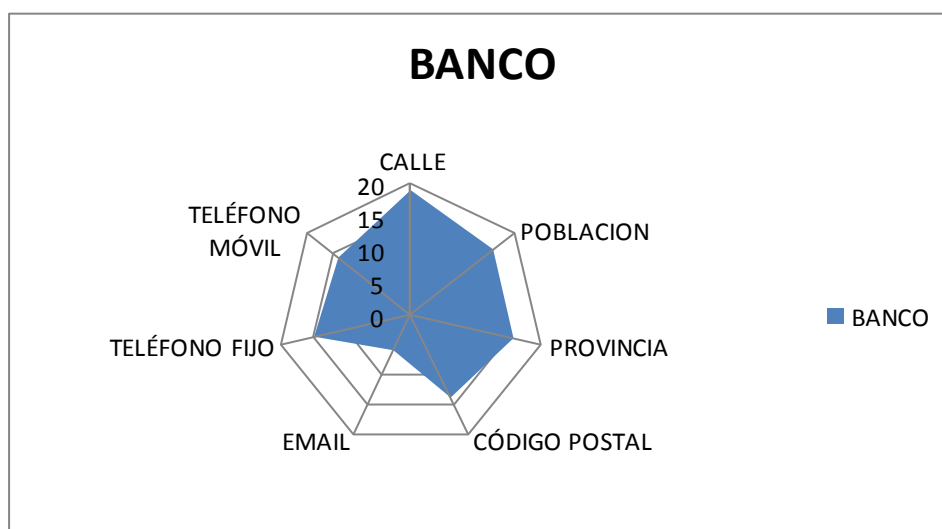


Ilustración 28. Gráfico de dirección de Banco



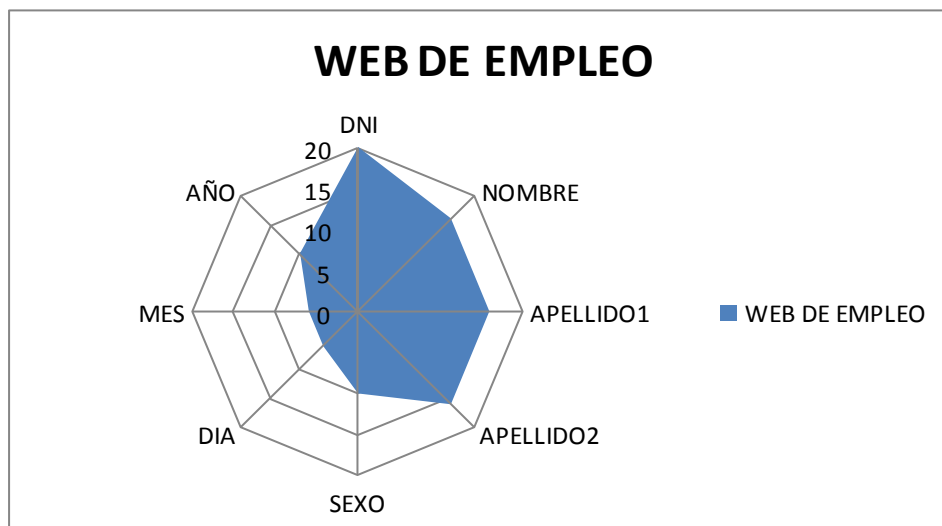
## Auditoría de calidad de bases de datos de carácter personal

- Si el usuario desea introducir los campos para su utilización en una base de datos de una Web de empleo, su ponderación será la siguiente:

Datos de identificación:

DNI	20
NOMBRE	16
APELLIDO1	16
APELLIDO2	16
SEXO	10
DIA	6
MES	6
AÑO	10

**Ilustración 29. Campos de identificación de Web de empleo**

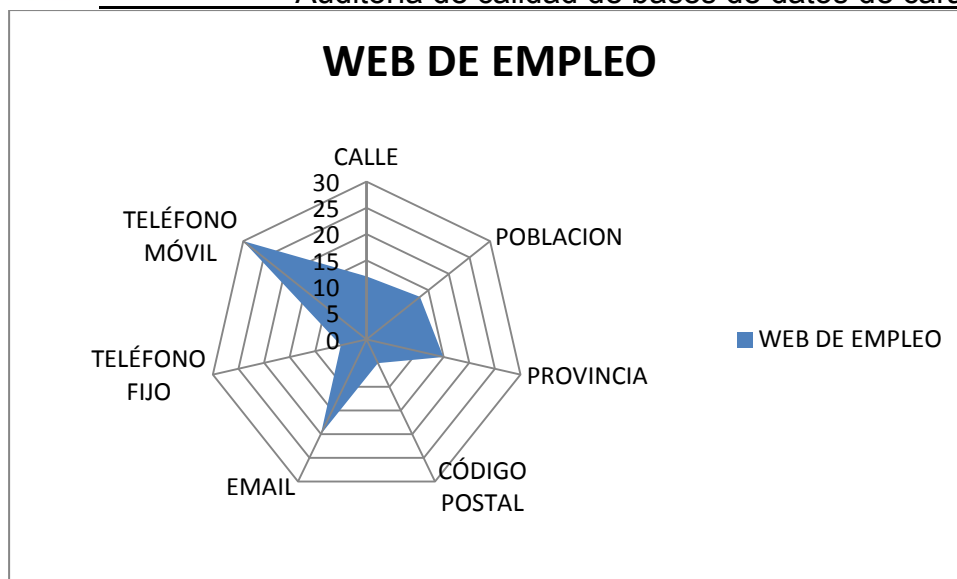


**Ilustración 30. Gráfico de identificación de Web de empleo**

Datos de dirección:

CALLE	12
POBLACION	13
PROVINCIA	15
CÓDIGO POSTAL	5
EMAIL	20
TELÉFONO FIJO	5
TELÉFONO MÓVIL	30

**Ilustración 31. Campos de dirección de Web de empleo**



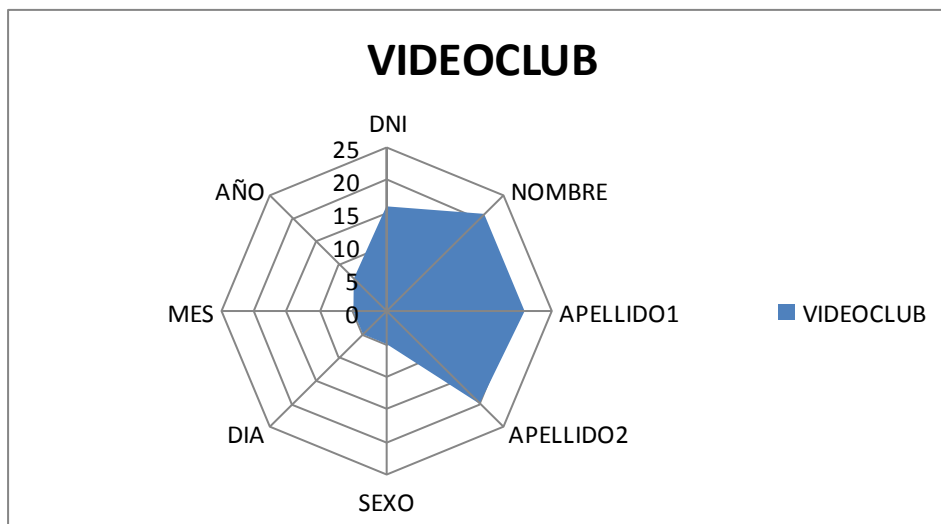
**Ilustración 32. Gráfico de dirección de Web de empleo**

- Por último, si el usuario desea introducir los campos para su uso en un Videoclub, éstos tendrán la siguiente ponderación:

Datos de identificación:

DNI	16
NOMBRE	21
APELLIDO1	21
APELLIDO2	20
SEXO	5
DIA	5
MES	5
AÑO	7

**Ilustración 33. Campos de identificación de Videoclub**

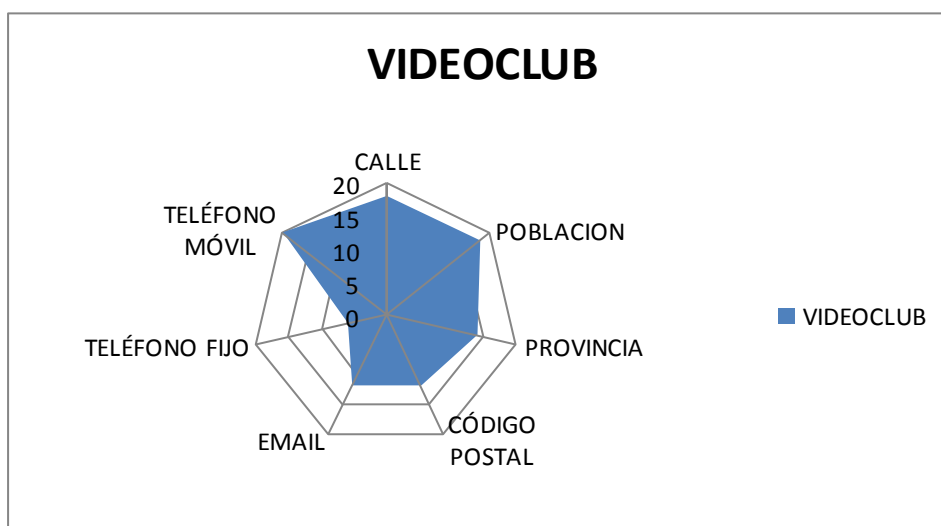


**Ilustración 34: Gráfico de identificación de Videoclub**

Datos de dirección:

CALLE	18
POBLACION	18
PROVINCIA	14
CÓDIGO POSTAL	12
EMAIL	12
TELÉFONO FIJO	6
TELÉFONO MÓVIL	20

**Ilustración 35. Campos de dirección de Videoclub**



**Ilustración 36. Gráfico de dirección de Videoclub**

(Existe una breve explicación del porqué de estas ponderaciones en el [Anexo](#))



#### **8.4 Penalizaciones a los resultados de la métrica.**

Antes de mostrar los resultados de la métrica al usuario, ésta comprobará si debe aplicar una serie de penalizaciones por los registros introducidos por el usuario (en forma de tanto por ciento).

- Si el usuario introduce los campos referentes a la identificación, una posible penalización estaría dada por falta de correspondencia del nombre con el sexo, es decir, una mujer, por ejemplo, no puede tener el nombre “Pablo”, por lo que dañará el resultado de las ponderaciones realizadas con anterioridad por el sistema.
- Si el usuario introduce los campos referentes a la identificación, otra posible penalización estaría dada por falta de correspondencia de un mes con el número de día, es decir, el mes “Febrero”, por ejemplo, no puede contener “30” días.
- Si el usuario introduce los campos referentes a la dirección, una posible penalización estaría dada por la falta de correspondencia entre la provincia y la población, es decir, pueden existir los dos por separado, pero no corresponder. Por ejemplo, “Cuenca” y “Colmenarejo” existen por separado, pero no corresponden.
- Si el usuario introduce los campos referentes a la dirección, otra posible penalización estaría dada por la falta de correspondencia entre la población y el código postal, es decir, pueden existir los dos por separado, pero no corresponder. Por ejemplo, “Collado-Villalba” y “29689” existen por separado, pero no corresponden.

Todas estas penalizaciones, penalizarán el resultado expresado por la métrica en un 15% en cada bbdd en la que se quiera insertar el registro solicitado por el usuario.

(Si con este 15% de penalización el registro llega a tener una equivalencia de “pésima” o “insuficiente” no se insertará tampoco en la bbdd correspondiente).



# Aplicación

Como ya hemos visto, existen numerosas empresas que se dedican a la normalización de las bases de datos de las empresas, cuya función es sanear las bases de datos existentes, aunque muy pocas, se centran en evaluar la calidad de los datos previa inserción en sus bases de datos.

Pues bien, a continuación se plantea el uso de una herramienta, de apoyo al auditor, para realizar una auditoría de calidad, de los datos que se van a introducir en diferentes bases de datos. Es decir, mediremos la salud con la que se van a introducir una serie de datos en diferentes bases de datos.

Para complementar esta herramienta se propone el uso de un pequeño checklist de apoyo al auditor, pues se pretende evaluar no solamente la función, sino también el contexto, dado que hablar de calidad, es también hablar de sus procedimientos de trabajo e instrucciones operativas, sus aspectos legales, sus aspectos administrativos, y, por último, de sus aspectos técnicos.



## 9.1 Checklist

Empezaremos por realizar una pequeña definición de la herramienta de apoyo checklist:

*“Un checklist, o lista de verificación, es un documento que detalla uno por uno distintos aspectos que se deben analizar, comprobar, verificar, etc.*

*Existen dos tipos fundamentales de checklist:*

- *De Rango: Preguntas o conceptos a evaluar en un rango determinado (permiten mayor precisión si el criterio de la Auditoría es uniforme). Por ejemplo de 0 a 10.*
- *Binarias: Preguntas o conceptos con respuesta única y excluyente, Si o No, 1 ó 0 (son excelentes si los cuestionarios están muy cuidados en su formulación).”*

El checklist implementado para ayudar al auditor en la tarea de medir la calidad de los datos, llevará una pequeña medición, para que el auditor pueda, después de rellenar dicho checklist, saber con qué nivel de calidad está implementado el formulario a rellenar para introducir los datos en la base de datos, e ir con ese dato a la tabla de equivalencias de la métrica diseñada en el apartado anterior, y poder así obtener un rango equivalente (aunque en este caso se hace caso omiso al umbral de calidad).

El checklist para el auditor constará de 27 verificaciones, las cuales irán divididas en cuatro partes diferenciadas, en las que ponderará de una manera u otra el nivel de calidad del mismo:

- Tipo-1 (Binario). Procedimientos de trabajo e instrucciones operativas:
  - En este apartado las verificaciones que no se cumplan restarán un 3,5% y las positivas sumarán otro 3,5%.
- Tipo-2 (Binario). Aspectos administrativos.
  - En este apartado las verificaciones que no se cumplan restarán un 4% y las positivas sumarán otro 4%.
- Tipo-3 (Binario). Aspectos legales.
  - En este apartado las respuestas positivas sumarán un 6%. Es en este punto cuando habrá alguna verificación que, en caso de no cumplirse, será definitiva para no autorizar la puesta en marcha del formulario.
- Tipo-4 (De rango). Aspectos técnicos.
  - En este apartado las ponderaciones tendrán un rango de 1 a 5, donde el 1 mostrará la mayor disconformidad del usuario en términos de calidad, y el 5 el máximo. Por lo cual una puntuación de “1”, restará un 3%, una puntuación de “2”, restará un 2%, una puntuación de “3”, sumará un 1%, una puntuación de “4”, sumará un 2%, y, por último, una puntuación de “5”, sumará un 3%.



Cabe destacar que el checklist, de acuerdo al reglamento de la AEPD de medidas de seguridad de los ficheros que contienen datos de carácter personal, cubrirá el nivel básico, pues es el nivel que define los ficheros que contienen datos de carácter personal. Además, podemos comprobar que los campos expuestos en el formulario no contienen datos sobre la ideología, religión, creencias, origen racial, salud o vida sexual de las personas, así como datos relativos a infracciones administrativas o penales de las personas, en cuyo caso el nivel de seguridad sería medio o incluso alto (algunas de estas medidas de seguridad están descritas en el artículo 5 del *Reglamento de Desarrollo de la Ley Orgánica 15/1999 de 13 de diciembre, de protección de datos de carácter personal*).

El resultado de todas las verificaciones se encuadrarán en dos apartados, *éxito* y *fracaso*.

- Éxito: resultará del sumatorio de todas las verificaciones positivas, en el caso de las binarias, y el sumatorio de las verificaciones ponderas en “3”, “4” y “5” en el caso de las verificaciones de rango.
- Fracaso: resultará del sumatorio de todas las verificaciones negativas, en el caso de las binarias, y el sumatorio de las verificaciones ponderas en “1” y “2” en el caso de las verificaciones de rango.

Así pues, el checklist será el siguiente:

### *Lista de verificaciones:*

*SI NO N/A*

#### **Tipo-1: Procedimientos de trabajo e instrucciones operativas**

¿Se tiene que usar un software especial para rellenar el formulario? (*)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
¿Existe un TestPlan (plan de pruebas) cada vez que se modifica el formulario?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
¿Se realizan revisiones periódicas sobre el sistema diseñado por parte del propietario del mismo?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
¿Existe algún tipo de formación para el personal que intervenga en la auditoría del formulario?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
¿Existe algún procedimiento para la realización de copias de respaldo y recuperación de datos?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
¿Existe algún procedimiento para la respuesta ante incidencias?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>





(\*) Esta verificación pondera al revés que las verificaciones de su grupo, es decir, si resulta positiva se restará un 3,5% sobre el total y si resulta negativa se sumará un 3,5% sobre el total de calidad.

*Lista de verificaciones:*

*SI NO N/A*

**Tipo-2: Aspectos administrativos**

¿Existe un administrador de bases de datos que gestione las instancias de las bases de datos?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
¿Se sabe quién es el propietario del formulario?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
¿Está disponible el código fuente?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
¿Se realizan copias de seguridad del formulario?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
¿Se indica al usuario la versión del formulario actual?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
¿Existe una gestión de incidencias (si presenta alguna incidencia el formulario)?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
¿Existen procedimientos de identificación y autenticación?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**Tipo-3: Aspectos legales**

¿Está dada de alta la base de datos en la AGPD?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
¿Existen contratos de confidencialidad para el usuario?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
¿Se muestra como el usuario puede ejercitar sus derechos de acceso, rectificación, cancelación y oposición (A.R.C.O) al fichero que contiene sus datos personales, según la LOPD?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



*Lista de verificaciones:*

1 2 3 4 5

**Tipo-4: Aspectos técnicos**

¿Son correctos los procedimientos internos para la operación de modificación del fichero?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
¿Son correctos los procedimientos internos para la actualización del fichero?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
¿Son correctos los procedimientos para avisar a alguien si hay algún fallo o problema al rellenar el formulario?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
¿El administrador ha realizado un ranking de los elementos más críticos?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
¿Son correctos el diseño físico y lógico de la base de datos?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
¿Existen criterios de acceso en relación a procedimientos de asignación y gestión de contraseñas?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
¿Son claros los campos a rellenar? ¿Se aprecia la diferencia entre ellos?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
¿Resulta extremadamente corto el formulario a rellenar?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
¿Resulta extremadamente largo el formulario a rellenar?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
¿Se obliga al relleno de todos los campos del formulario?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
¿Los campos a rellenar del formulario son reales?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Una vez el auditor rellene este checklist, podrá medir la calidad con la que se encuentra el formulario de entrada a la base de datos, y consultar la tabla de equivalencias de la métrica para saber en cual se encuentra el nivel de salud medido.



Un ejemplo del checklist rellenado por el auditor sería el siguiente:

<i>Lista de verificaciones:</i>	SI	NO	N/A
<b>Tipo-1: Procedimientos de trabajo e instrucciones operativas</b>			
¿Se tiene que usar un software especial para rellenar el formulario?		X	
¿Existe un TestPlan (plan de pruebas) cada vez que se modifica el formulario?		X	
¿Se realizan revisiones periódicas sobre el sistema diseñado por parte del propietario del mismo?		X	
¿Existe algún tipo de formación para el personal que intervenga en la auditoría del formulario?	X		
¿Existe algún procedimiento para la realización de copias de respaldo y recuperación de datos?	X		
¿Existe algún procedimiento para la respuesta ante incidencias?	X		
<b>Tipo-2: Aspectos administrativos</b>			
¿Existe un administrador de bases de datos que gestione las instancias de las bases de datos?	X		
¿Se sabe quién es el propietario del formulario?		X	
¿Está disponible el código fuente?	X		
¿Se realizan copias de seguridad del formulario?	X		
¿Se indica al usuario la versión del formulario actual?		X	
¿Existe una gestión de incidencias (si presenta alguna incidencia el formulario)?	X		
¿Existen procedimientos de identificación y autenticación?		X	
<b>Tipo-3: Aspectos legales</b>			
¿Está dada de alta la base de datos en la AGPD?	X		
¿Existen contratos de confidencialidad para el usuario?	X		
¿Se muestra como el usuario puede ejercitar sus derechos de acceso, rectificación, cancelación y oposición (A.R.C.O) al fichero que contiene sus datos personales, según la LOPD?	X		

Tabla 2. Ejemplo de checklist\_1



Lista de verificaciones:					
	1	2	3	4	5
<b>Tipo-4: Aspectos técnicos</b>					
¿Son correctos los procedimientos internos para la operación de modificación del fichero?			X		
¿Son correctos los procedimientos internos para la actualización del fichero?			X		
¿Son correctos los procedimientos para avisar a alguien si hay algún fallo o problema al rellenar el formulario?			X		
¿El administrador ha realizado un ranking de los elementos más críticos?	X				
¿Existe diseño físico y lógico de la base de datos?				X	
¿Existen criterios de acceso en relación a procedimientos de asignación y gestión de contraseñas?			X		
¿Son claros los campos a rellenar? ¿Se aprecia la diferencia entre ellos?					X
¿Resulta extremadamente corto el formulario a rellenar ?				X	
¿Resulta extremadamente largo el formulario a rellenar?				X	
¿Se obliga al relleno de todos los campos del formulario?			X		
¿Los campos a rellenar del formulario son reales?					X

Tabla 3. Ejemplo de checklist\_2

Tras el cual, la medición de la salud del formulario y de la base de datos sería la siguiente:

NIVEL DE SALUD DEL FORMULARIO DE LA BBDD	
ÉXITO	65%
FRACASO	22%

Tabla 4. Nivel de salud del formulario

$$(\text{Éxito} = (4 * 3,5\%) + (4 * 4\%) + (3 * 6\%) + (5 * 1\%) + (3 * 2\%) + (2 * 3\%) = 65\%)$$

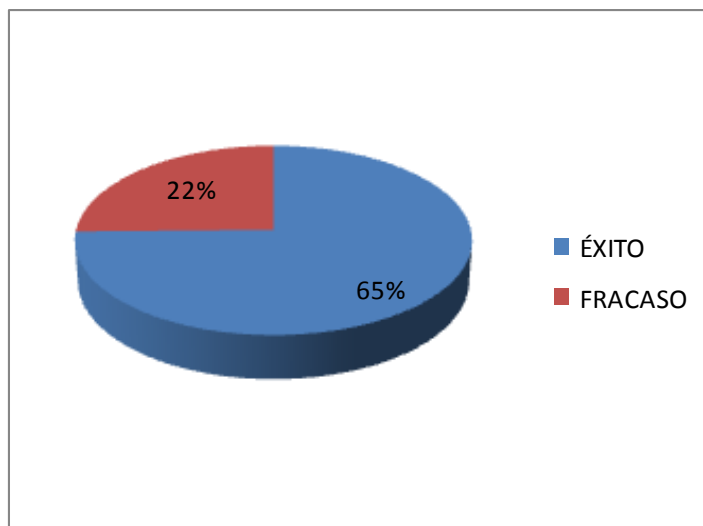
$$(\text{Fracaso} = (2 * 3,5\%) + (3 * 4\%) + (1 * 3\%) = 22\%)$$



### Auditoría de calidad de bases de datos de carácter personal

El resultado de la suma de los dos apartados no resulta un 100% pues hay verificaciones (en el grupo de las verificaciones de rango) que no tienen porqué sumar o restar el máximo de su valor (si el usuario las pondera con un “2”, un “3” o un “4”).

Llevada a un pequeño gráfico de sectores (ponderado sobre el 87% alcanzado, no sobre el 100%):



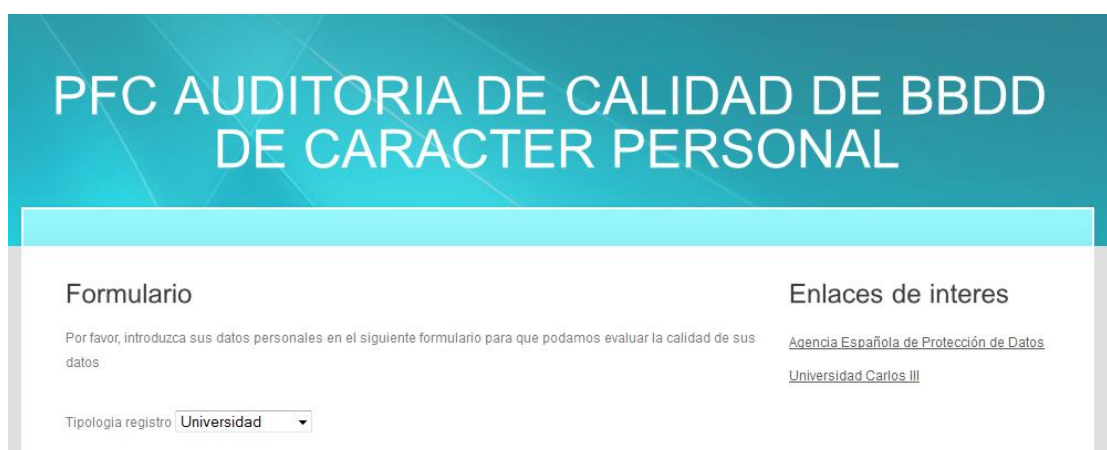
**Ilustración 37. Gráfico del nivel de salud del formulario**

Como se puede observar, en este ejemplo implementado, el nivel de salud corresponde a un 65% sobre el total, el cual, yendo a la tabla de equivalencias de mediciones para la métrica elaborada, explicada en la introducción de este apartado, mostraría un nivel de calidad “Bueno”.

## 9.2 Servicio Web

Paralelo al Checklist implementado, se plantea el uso de un servicio web en el que el auditor podrá evaluar la calidad de los datos que se van a insertar en cada una de las bases de datos diseñadas, mediante la inserción de los registros que contendrán cada uno de los campos de carácter personal en un pequeño formulario.

El formulario posee la siguiente interfaz inicial:



**PFC AUDITORIA DE CALIDAD DE BBDD DE CARACTER PERSONAL**

**Formulario**

Por favor, introduzca sus datos personales en el siguiente formulario para que podamos evaluar la calidad de sus datos

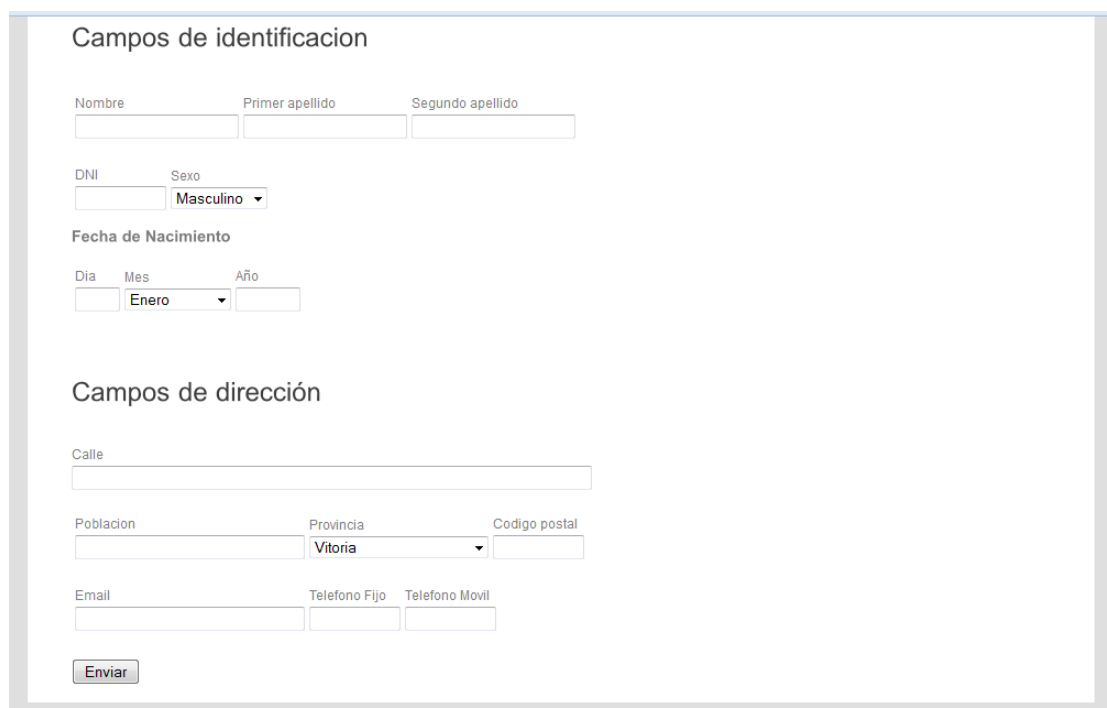
Tipologia registro:

**Enlaces de interes**

[Agencia Española de Protección de Datos](#)

[Universidad Carlos III](#)

Ilustración 38. Interfaz principal parte 1



**Campos de identificación**

Nombre  Primer apellido  Segundo apellido

DNI  Sexo

Fecha de Nacimiento

Día  Mes  Año

**Campos de dirección**

Calle

Poblacion  Provincia  Código postal

Email  Telefono Fijo  Telefono Movil

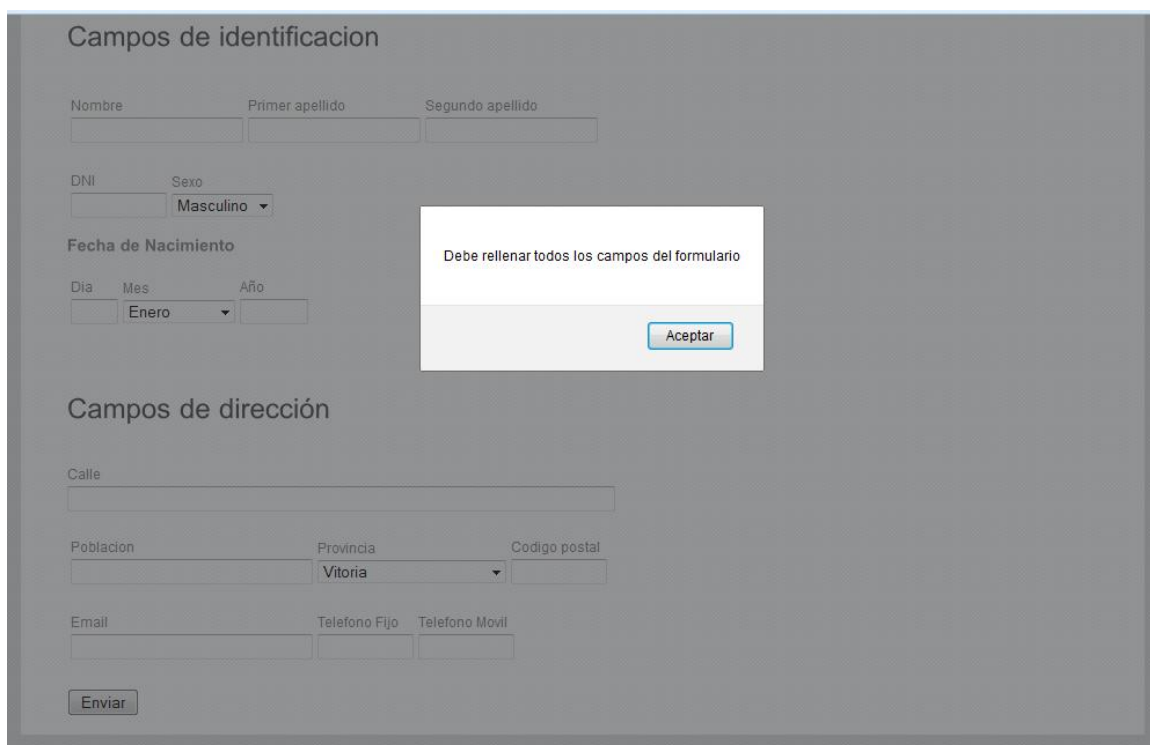
Ilustración 39. Interfaz principal parte 2

En la que podemos observar cómo están todos los campos descritos con anterioridad, separados por las agrupaciones de identificación y de dirección.

### 9.2.1 Pruebas Formulario

Para probar el formulario implementado se ha procedido a realizar infinidad de pruebas, para observar su grado de fiabilidad, aunque para poder evidenciarlo de un modo más genérico se han agrupado en ocho pruebas generales (todas ellas para un registro con la tipología de “Universidad”):

1. El usuario envía el formulario completamente en blanco:



Campos de identificación

Nombre  Primer apellido  Segundo apellido

DNI  Sexo

Fecha de Nacimiento

Día  Mes  Año

Debe rellenar todos los campos del formulario

Aceptar

Campos de dirección

Calle

Poblacion  Provincia  Codigo postal

Email  Telefono Fijo  Telefono Movil

Enviar

Ilustración 40. Ejemplo formulario 1



2. El usuario envía el formulario con algún campo en blanco:

The screenshot shows a web form titled 'Campos de identificación' and 'Campos de dirección'. The form contains several input fields for personal data. A modal dialog box is displayed in the center, indicating a validation error: 'Debe rellenar todos los campos del formulario' (You must fill out all the fields of the form). The dialog has an 'Aceptar' (Accept) button. The form fields include:

- Campos de identificación:**
  - Nombre: Alejandro
  - Primer apellido: (empty)
  - Segundo apellido: Lara
  - DNI: 70073728L
  - Sexo: Masculino (dropdown)
  - Fecha de Nacimiento: Dia 11, Mes Noviembre (dropdown), Año 1986
- Campos de dirección:**
  - Calle: (empty)
  - Población: Collado Villalba
  - Provincia: Madrid (dropdown)
  - Código postal: 28400
  - Email: (empty)
  - Telefono Fijo: (empty)
  - Telefono Movil: (empty)

An 'Enviar' (Send) button is located at the bottom left of the form.

Ilustración 41. Ejemplo formulario 2





### Auditoría de calidad de bases de datos de carácter personal

3. El usuario envía el formulario rellenado completamente, con diversos errores en los campos de identificación y dirección (por debajo del umbral de calidad):

**Campos de identificación**

Nombre	Primer apellido	Segundo apellido
<input type="text" value="Alejandro"/>	<input type="text" value="Viana"/>	<input type="text" value="Laara"/>

DNI	Sexo
<input type="text" value="70073729L"/>	<input type="text" value="Masculino"/>

Fecha de Nacimiento

Día	Mes	Año
<input type="text" value="ee"/>	<input type="text" value="Enero"/>	<input type="text" value="eeee"/>

**Campos de dirección**

Calle
<input type="text" value="88"/>

Poblacion	Provincia	Codigo postal
<input type="text" value="Colme"/>	<input type="text" value="Vitoria"/>	<input type="text" value="eeee"/>

Email	Telefono Fijo	Telefono Movil
<input type="text" value="ewepeqwewe"/>	<input type="text" value="eeeeeeee"/>	<input type="text" value="0e191"/>

**Ilustración 42. Ejemplo formulario 3**



Cuyos resultados son los que se muestran a continuación:

## Resultados

**El registro no es apto para introducirse en la base de datos**

Tipología del registro: Universidad

- Nombre: Aljeandro ✖
- Primer apellido: Viana ✖
- Segundo apellido: Laara ✖
- DNI: 70073729L ✖
- Sexo: Masculino ✔
- Día de nacimiento: ee ✖
- Mes de nacimiento: Enero ✔
- Año de nacimiento: eeee ✖
- Calle: 88 ✔
- Población: Colme ✖
- Código postal: eeeee ✖
- Provincia: Vitoria ✔
- Email: eweqeqwewe ✖
- Teléfono fijo: eeeeeeee ✖
- Teléfono móvil: eeeee ✖

### Ilustración 43. Resultado ejemplo formulario 3 - 1

- Penalización por nombre que no corresponde al sexo: ✖
- Penalización por día de nacimiento no se corresponde con el mes de nacimiento: ✖
- Penalización por población que no corresponde con la provincia: ✖
- Penalización por código postal no corresponde con población: ✖

Valoración identificación: 17

Valoración dirección: 30

**Valoración Total: 22,2**

### Ilustración 44. Resultado ejemplo formulario 3 - 2



4. El usuario envía el formulario rellenado completamente, con diversos errores en los campos de identificación y dirección (por encima del umbral de calidad):

### Campos de identificación

Nombre: Alejandro  
Primer apellido: Viiana  
Segundo apellido: Lara

DNI: 70073729L  
Sexo: Masculino

### Fecha de Nacimiento

Día: 11  
Mes: Enero  
Año: 1986

### Campos de dirección

Calle: 88

Poblacion: Collado Villalba  
Provincia: Madrid  
Codigo postal: eeeee

Email: eweqeqwewe  
Telefono Fijo: 678903315  
Telefono Movil: eeeeeee

Enviar

Ilustración 45. Ejemplo formulario 4



Cuyos resultados son los que se muestran a continuación:

## Resultados

**El registro es apto para introducirse en la base de datos**

Tipología del registro: Universidad

- Nombre: Alejandro ✓
- Primer apellido: Viana ✗
- Segundo apellido: Lara ✓
- DNI: 70073729L ✗
- Sexo: Masculino ✓
- Día de nacimiento: 11 ✓
- Mes de nacimiento: Enero ✓
- Año de nacimiento: 1986 ✓
- Calle: 88 ✓
- Población: Collado Villalba ✓
- Código postal: eeeee ✗
- Provincia: Madrid ✓
- Email: eweqeqwewe ✗
- Teléfono fijo: 678903315 ✓
- Teléfono móvil: eeeee ✗

### Ilustración 46. Resultado ejemplo formulario 4 - 1

- Penalización por nombre que no corresponde al sexo: ✗
- Penalización por día de nacimiento no se corresponde con el mes de nacimiento: ✗
- Penalización por población que no corresponde con la provincia: ✗
- Penalización por código postal no corresponde con población: ✗

Valoración identificación: 68

Valoración dirección: 45

**Valoración Total: 58,8**

### Ilustración 47. Resultado ejemplo formulario 4 - 2



## Auditoría de calidad de bases de datos de carácter personal

5. El usuario envía el formulario rellenado completamente, sin ningún error en sus campos, aunque con penalizaciones de identificación a aplicar:

### Campos de identificación

Nombre	Primer apellido	Segundo apellido
<input type="text" value="Laura"/>	<input type="text" value="Viana"/>	<input type="text" value="Lara"/>
DNI	Sexo	
<input type="text" value="70073728L"/>	<input type="text" value="Masculino"/>	
Fecha de Nacimiento		
Día	Mes	Año
<input type="text" value="11"/>	<input type="text" value="Enero"/>	<input type="text" value="1986"/>

### Campos de dirección

Calle		
<input type="text" value="Montenebros II"/>		
Poblacion	Provincia	Codigo postal
<input type="text" value="Collado Villalba"/>	<input type="text" value="Madrid"/>	<input type="text" value="28400"/>
Email	Telefono Fijo	Telefono Movil
<input type="text" value="avianalara@hotmail.com"/>	<input type="text" value="678903315"/>	<input type="text" value="918508293"/>
<input type="button" value="Enviar"/>		

**Ilustración 48. Ejemplo formulario 5**



Cuyos resultados son los que se muestran a continuación:

## Resultados

El registro es apto para introducirse en la base de datos

Tipología del registro: Universidad

- Nombre: Laura ✓
- Primer apellido: Viana ✓
- Segundo apellido: Lara ✓
- DNI: 70073728L ✓
- Sexo: Masculino ✓
- Día de nacimiento: 11 ✓
- Mes de nacimiento: Enero ✓
- Año de nacimiento: 1986 ✓
- Calle: Montenebros II ✓
- Población: Collado Villalba ✓
- Código postal: 28400 ✓
- Provincia: Madrid ✓
- Email: avianalara@hotmail.com ✓
- Teléfono fijo: 678903315 ✓
- Teléfono móvil: 918508293 ✓

### Ilustración 49. Resultado ejemplo formulario 5 - 1

- Penalización por nombre que no corresponde al sexo: ✓
- Penalización por día de nacimiento no se corresponde con el mes de nacimiento: ✗
- Penalización por población que no corresponde con la provincia: ✗
- Penalización por código postal no corresponde con población: ✗

Valoración identificación: 85

Valoración dirección: 100

**Valoración Total: 91**

### Ilustración 50. Resultado ejemplo formulario 5 - 2



6. El usuario envía el formulario rellenado completamente, sin ningún error en sus campos, aunque con penalizaciones de dirección a aplicar:

---

### Campos de identificación

Nombre	Primer apellido	Segundo apellido
<input type="text" value="Alejandro"/>	<input type="text" value="Viana"/>	<input type="text" value="Lara"/>

DNI	Sexo
<input type="text" value="70073728L"/>	<input type="text" value="Masculino"/>

Fecha de Nacimiento

Día	Mes	Año
<input type="text" value="11"/>	<input type="text" value="Enero"/>	<input type="text" value="1986"/>

### Campos de dirección

Calle
<input type="text" value="Montenebros II"/>

Poblacion	Provincia	Codigo postal
<input type="text" value="Collado Villalba"/>	<input type="text" value="Murcia"/>	<input type="text" value="28400"/>

Email	Telefono Fijo	Telefono Movil
<input type="text" value="avianalara@hotmail.com"/>	<input type="text" value="678903315"/>	<input type="text" value="918508293"/>

---

Ilustración 51. Ejemplo formulario 6



Cuyos resultados son los que se muestran a continuación:

## Resultados

El registro es apto para introducirse en la base de datos

Tipología del registro: Universidad

- Nombre: Alejandro ✓
- Primer apellido: Viana ✓
- Segundo apellido: Lara ✓
- DNI: 70073728L ✓
- Sexo: Masculino ✓
- Día de nacimiento: 11 ✓
- Mes de nacimiento: Enero ✓
- Año de nacimiento: 1986 ✓
- Calle: Montenebros II ✓
- Población: Collado Villalba ✓
- Código postal: 28400 ✓
- Provincia: Murcia ✓
- Email: avianalara@hotmail.com ✓
- Teléfono fijo: 678903315 ✓
- Teléfono móvil: 918508293 ✓

### Ilustración 52. Resultado ejemplo formulario 6 - 1

- Penalización por nombre que no corresponde al sexo: ✗
- Penalización por día de nacimiento no se corresponde con el mes de nacimiento: ✗
- Penalización por población que no corresponde con la provincia: ✓
- Penalización por código postal no corresponde con población: ✗

Valoración identificación: 100

Valoración dirección: 85

**Valoración Total: 94**

### Ilustración 53. Resultado ejemplo formulario 6 – 2





7. El usuario envía el formulario relleno completamente sin ningún error en sus campos y ninguna penalización a aplicar:

---

### Campos de identificación

Nombre	Primer apellido	Segundo apellido
<input type="text" value="Alejandro"/>	<input type="text" value="Viana"/>	<input type="text" value="Lara"/>

DNI	Sexo
<input type="text" value="70073728L"/>	<input type="text" value="Masculino"/>

### Fecha de Nacimiento

Día	Mes	Año
<input type="text" value="11"/>	<input type="text" value="Enero"/>	<input type="text" value="1986"/>

### Campos de dirección

Calle
<input type="text" value="Montenebros II"/>

Poblacion	Provincia	Codigo postal
<input type="text" value="Collado Villalba"/>	<input type="text" value="Madrid"/>	<input type="text" value="28400"/>

Email	Telefono Fijo	Telefono Movil
<input type="text" value="avianalara@hotmail.com"/>	<input type="text" value="678903315"/>	<input type="text" value="918508293"/>

---

**Ilustración 54. Ejemplo formulario 7**



Cuyos resultados son los que se muestran a continuación:

## Resultados

El registro es apto para introducirse en la base de datos

Tipología del registro: Universidad

- Nombre: Alejandro ✓
- Primer apellido: Viana ✓
- Segundo apellido: Lara ✓
- DNI: 70073728L ✓
- Sexo: Masculino ✓
- Día de nacimiento: 11 ✓
- Mes de nacimiento: Enero ✓
- Año de nacimiento: 1986 ✓
- Calle: Montenebros II ✓
- Población: Collado Villalba ✓
- Código postal: 28400 ✓
- Provincia: Madrid ✓
- Email: avianalara@hotmail.com ✓
- Teléfono fijo: 678903315 ✓
- Teléfono móvil: 918508293 ✓

### Ilustración 55. Resultado ejemplo formulario 7 - 1

- Penalización por nombre que no corresponde al sexo: ✗
- Penalización por día de nacimiento no se corresponde con el mes de nacimiento: ✗
- Penalización por población que no corresponde con la provincia: ✗
- Penalización por código postal no corresponde con población: ✗

Valoración identificación: 100

Valoración dirección: 100

**Valoración Total: 100**

### Ilustración 56. Resultado ejemplo formulario 7 - 2



# Bloque IV

## Presupuesto



UNIVERSIDAD CARLOS III DE MADRID  
Escuela Politécnica Superior

PRESUPUESTO DE  
PROYECTO

1.- Autor: Alejandro Viana Lara  
2.- Departamento: Informática, Auditoría Informática  
3.- Descripción del Proyecto:

- Título Auditoría de calidad de bases de datos de carácter personal  
- Duración (meses) 5  
Tasa de costes Indirectos: N/A (debido a las características de este Proyecto)

4.- Presupuesto total del Proyecto (valores en Euros):  
12933,07 Euros

5.- Desglose presupuestario  
(costes directos) PERSONAL

Apellidos y nombre	N.I.F. (no rellenar - solo a título)	Categoría	Dedicación (hombres mes)	Coste hombre mes	Coste (Euro)	Firma de conformidad
Alejandro Viana Lara		Ingeniero técnico	5	2.694,39	13471,95	
Hombres mes 5				Total	13471,95	

a) 1 Hombre mes = 131,25 horas. Máximo anual de dedicación de 12 hombres mes (1575 horas) Máximo anual para PDI de la Universidad Carlos III de Madrid de 8,8 hombres mes (1.155 horas)

### EQUIPO

Descripción	Coste (Euro)	% Uso dedicado proyecto	Dedicación (meses)	Periodo de depreciación <sup>d)</sup>	Coste imputable <sup>d)</sup>
		100		60	0,00
		100		60	0,00
		100		60	0,00
		100		60	0,00
		100		60	0,00
Total					0,00

d) Fórmula de cálculo de la Amortización:

**A**  
**x C x**  
**D B**

**A** = nº de meses desde la fecha de facturación en que el equipo es utilizado  
**B** = periodo de depreciación (60 meses)  
**C** = coste del equipo (sin IVA)  
**D** = % del uso que se dedica al proyecto (habitualmente 100%)



## Auditoría de calidad de bases de datos de carácter personal

### SUBCONTRATACIÓN DE TAREAS

Descripción	Empresa	Coste imputable
		0,00

### OTROS COSTES DIRECTOS DEL PROYECTO<sup>e)</sup>

Descripción	Empresa	Costes imputable
		0,00

e) Este capítulo de gastos incluye todos los gastos no contemplados en los conceptos anteriores, por ejemplo: fungible, viajes y dietas, otros,...

### 6 - Resumen de costes

Presupuesto Costes Totales	Presupuesto Costes Totales
Personal	13471,95
Amortización	0*
Subcontratación de tareas	0
Costes de funcionamiento	0
Costes Indirectos	0
Total	13471,95

\* Las amortizaciones resultan de un coste cero, pues las licencias del software utilizado han sido extraídas del MSN Academy de Microsoft, para el cual la Universidad posee un contrato con ellos, y el portátil utilizado ya fue costado con anterioridad al desarrollo de este proyecto.

“El presupuesto total de este proyecto asciende a la cantidad de 13471,95 EUROS.

Colmenarejo a 15 de enero de 2013

El ingeniero proyectista

Fdo. Alejandro Viana Lara”



# Líneas de investigación futuras

Como en todo trabajo de análisis y desarrollo, a medida que se avanza en el estudio surgen nuevas vías o entornos a estudiar que pueden mejorar e incrementar la calidad de nuestro estudio inicial.

Por ello, este proyecto también nace con la pretensión de poder servir de punto de partida para posibles futuros proyectos de fin de carrera, relacionados con la auditoría de calidad de las bases de datos.

En este proyecto se expone una herramienta de apoyo al auditor para evaluar el nivel de salud de los datos a introducir en diferentes bases de datos.

Una herramienta para complementarla podría ser una aplicación que evaluase el nivel de salud de los datos de una base de datos existente, para que la persona encargada de introducir registros en la base de datos correspondiente, supiera, antes de introducir nuevos datos, la calidad con la que cuentan los datos de la misma.

Aunque, la mayor herramienta a complementar este proyecto sería un prototipo que tras indicar el nivel de salud de los datos a introducir, indicara para los diferentes “errores” o “fallos en los campos” aproximaciones de los términos que pudieran resultar idóneos. Es decir, si el usuario introduce el campo *Aljeandro*, la herramienta además de iniciarle un ínfimo nivel de salud en el campo nombre, le indicara que quizá hubiera deseado introducir el campo *Alejandro*.

# Conclusiones

Como se ha podido comprobar en todo el estudio que ha envuelto a este proyecto, actualmente existe un problema mayúsculo en las bases de datos de nuestra sociedad, derivado de la falta de calidad de las mismas.

Es por ello por lo que las empresas para conseguir mejorar este aspecto, necesitan una serie de metodologías o pasos a seguir, que las permitan mejorar en cuanto a esa calidad tan deseada.

Al ser la calidad de datos un fin tan deseado ¿porqué no invertir en ella? Al realizar este Proyecto, se puede llegar a entender que este acto no es un gasto sin más, sino una gran inversión, pues el éxito de una empresa depende de un modo directo de la eficiencia de sus sistemas.

Resulta totalmente necesario e imprescindible que las empresas y organizaciones de hoy en día, tengan implantadas una serie de aplicaciones como la descrita en este Proyecto, con el fin de mantenerse al día en temas de calidad y así poder ofrecer un mejor servicio a sus respectivos clientes, además de reducir los efectos negativos en sus operaciones.

Cabe recalcar, siendo realistas, que no se puede lograr la calidad total, pero sí que se puede mejorar la calidad hasta el punto en el que la información sea adecuada para la toma de decisiones de una entidad, puesto que unos datos de calidad, implicarán un servicio de calidad.

Todo esto hace que considere alcanzado el objetivo que me había propuesto con el desarrollo de este Proyecto de Fin de Carrera, además de haber obtenido, o, por lo menos, eso pienso, un total enriquecimiento tanto a nivel técnico como a nivel personal.

# Anexo

## **Bibliografía**

A continuación se muestra la bibliografía utilizada en la elaboración de este Proyecto, según orden de aparición en el mismo, y desglosadas según la fuente utilizada:

[AEPD] *Agencia Española de Protección de Datos*. Disponible [Internet]: <http://www.agpd.es> [21 de diciembre de 2011]

[WIKI] *Wikipedia. The free Encyclopedia*. Disponible [Internet]: <http://www.wikipedia.org> [21 de diciembre de 2011]

[SCDMD] *Sigma. Calidad de Datos, motor de desarrollo de la empresa*. Disponible [Internet]: <http://marketing.sigma-data.com/noticias/4-noticias/11-calidad-de-datos-motor-de-desarrollo-de-la-empresa> [21 de diciembre de 2011]

[SCNCD] *Sigma. El coste de la no calidad de datos*. Disponible [Internet]: <http://marketing.sigma-data.com/noticias/4-noticias/17-el-coste-de-la-no-calidad-de-datos> [21 de diciembre de 2011]

[SCFNC] *Sigma. Las 10 causas más frecuentes de la no calidad de datos*. Disponible [Internet]: <http://marketing.sigma-data.com/noticias/4-noticias/14-las-10-causas-mas-frecuentes-de-la-no-calidad-de-datos> [21 de diciembre de 2011]

[ARTIC] *Articuloz. Factores de la calidad de datos*. Disponible [Internet]: <http://www.articuloz.com/empresas-articulos/la-importancia-de-la-auditoria-de-lacalidad-771130.html> [21 de diciembre de 2011]



- [MCD] *Microsoft. Calidad de datos.* Disponible [Internet]: [http://www.microsoft.com/business/smb/es-es/guia\\_lopd/calidad\\_datos.mspx](http://www.microsoft.com/business/smb/es-es/guia_lopd/calidad_datos.mspx) [7 de julio de 2011]
- [UPACD] *Universidad Pampa. Auditoría y control interno.* Disponible [Internet]: <http://www.eco.unlpam.edu.ar/Tteoricos/auditoria/Auditoria%20%20Calidad%20de%20Datos.pdf> [21 de diciembre de 2011]
- [LIAC] *La importancia de la auditoría de calidad.* Disponible [Internet]: <http://www.articuloz.com/empresas-articulos/la-importancia-de-la-auditoria-de-la-calidad-771130.html> [15 de febrero de 2009]
- [DEYDE] *Deyde, empresa dedicada a la normalización de nombres.* Disponible [Internet]: [www.deyde.es](http://www.deyde.es) [01 de diciembre de 2012]
- [CAMER] *Camerdata, empresa dedicada a las bases de datos internacionales.* Disponible [Internet]: [www.camerdata.es](http://www.camerdata.es) [01 de diciembre de 2012]
- [SCHOB] *Schober, empresa dedicada a las bases de datos de empresas.* Disponible [Internet]: [www.schober.es](http://www.schober.es) [01 de diciembre de 2012]
- [PITBO] *Pitney Bowes, empresa dedicada a ofrecer soluciones de calidad.* Disponible [Internet]: [www.pbinsight.com](http://www.pbinsight.com) [01 de diciembre de 2012]
- [AXESO] *Axesor, empresa dedicada al asesoramiento de normalizadores.* Disponible [Internet]: [www.axesor.es](http://www.axesor.es) [01 de diciembre de 2012]
- [TALEN] *Talend\*, empresa estadounidense dedicada a la calidad de los datos.* Disponible [Internet]: [www.talend.com/products/data-quality](http://www.talend.com/products/data-quality) [11 de diciembre de 2012]
- [BUDQU] *Business Data Quality, empresa londinense dedicada a la normalización de datos.* Disponible [Internet]: [www.businessdataquality.com](http://www.businessdataquality.com) [11 de diciembre de 2012]
- [DAQUO] *Data Quality Online, sitio web del gobierno australiano, dedicado a la medición de la calidad de los datos.* Disponible [Internet]: [www.nss.gov.au/dataquality](http://www.nss.gov.au/dataquality) [11 de diciembre de 2012]





[DATAF] *Dataflux, empresa estadounidense dedicada a la calidad de datos en general.* Disponible [Internet]: [www.dataflux.com/home.aspx?lang=es-es](http://www.dataflux.com/home.aspx?lang=es-es)  
[11 de diciembre de 2012]

[AIEP] Piattini Velthuis, Mario G. *Auditoría Informática: un enfoque práctico.* Segunda Edición. ISBN: 84-7897-293-5.

[IMD] Hernández Orallo, José. Isasi Viñuela, Pedro. *Introducción a la Minería de Datos.* Primera Edición. ISBN: 84-2054-091-9.

[FSBD] Elmasri, Ramez. Navathe, Shamkant. *Fundamentos de Sistemas de Bases de Datos.* Tercera Edición. ISBN: 84-7829-051-6.

[SGBD] Ramakrishnan, Raghu. Gehrke, Johannes. *Sistemas de Gestión de Bases de Datos.* Tercera Edición. ISBN: 0-07-246563-8.

[ASI] Bernal Montañés, Rafael. *Auditoría de los Sistemas de Información.* Primera Edición. ISBN: 84-7721-393-3.



## **Glosario de términos y acrónimos**

AEPD	<i>Agencia Española de Protección de Datos</i>
ARCO	<i>Derechos de acceso, rectificación, cancelación y oposición, a través de los cuales la LOPD garantiza a las personas el poder de control sobre sus datos personales</i>
Backup	<i>Copia de seguridad</i>
BBDD	<i>Base de Datos</i>
Checklist	<i>Lista de verificaciones</i>
CISA	<i>Certified Information Systems Auditor</i>
COBIT	<i>Control Objectives for Information and related Technology</i>
CRM	<i>Customer Relationship Management</i>
ISO	<i>International Organization for Standardization</i>
ISACA	<i>Information Systems Audit and Control Association</i>
LOPD	<i>Ley Orgánica de Protección de Datos</i>
LORTAD	<i>Ley Orgánica de Regulación del Tratamiento Automatizado de los Datos de Carácter Personal</i>
NOTA	<i>Notificaciones Telemáticas a la Agencia de Protección de Datos</i>
RAE	<i>Real Academia Española</i>
RGPD	<i>Registro General de Protección de Datos</i>
SSOO	<i>Sistema Operativo</i>
Web Service	<i>Servicio Web (software que utiliza un conjunto de protocolos y estándares que sirven para intercambiar datos entre aplicaciones)</i>
XML	<i>Extensible Markup Language</i>



## **Formulario NOTA en la AEPD**

### **¿Cómo se da de alta una base de datos, o un fichero en la AEPD? Formulario NOTA**

Al crear nuestra base de datos, tenemos que saber que siendo el contexto, datos de carácter personal, los datos se encuentran protegidos por la Ley Orgánica de Protección de Datos de Carácter Personal, más comúnmente conocida como la LOPD.

Dicho esto, tendremos que acudir a la Agencia Española de Protección de Datos, la cual dice lo siguiente:

*“El Registro General de Protección de Datos es el órgano de la Agencia Española de Protección de Datos al que corresponde velar por la publicidad de la existencia de los ficheros y tratamientos de datos de carácter personal, con miras a hacer posible el ejercicio de los derechos de información, acceso, rectificación y cancelación de datos regulados en los artículos 14 a 17 de la Ley Orgánica 15/99, de 13 de diciembre, de Protección de Datos de Carácter Personal.*

*Serán objeto de inscripción en el Registro General de Protección de Datos:*

- *Los ficheros de las Administraciones Públicas*
- *Los ficheros de titularidad privada*
- *Las autorizaciones de transferencias internacionales de datos de carácter personal con destino a países que no presten un nivel de protección equiparable al que presta la LOPD a que se refiere el art. 33.1 de la citada Ley.*
- *Los códigos tipo , a que se refiere el artículo 32 de la LOPD.*
- *Los datos relativos a los ficheros que sean necesarios para el ejercicio de los derechos de información, acceso, rectificación, cancelación y oposición.*

*Para realizar la inscripción inicial del fichero y, en su caso, la posterior modificación o supresión de la inscripción, se encuentra disponible el formulario electrónico NOTA (titularidad pública y titularidad privada) a través del que deberán efectuarse las solicitudes de inscripción de ficheros en el Registro General de Protección de Datos (aprobado mediante Resolución de la AEPD de 12 de julio de 2006- B.O.E. 181 de 31 de julio).*

*Este formulario permite la presentación de forma gratuita de notificaciones a través de Internet con certificado de firma electrónica. En caso de no disponer de un certificado de firma electrónica, también puede presentar la notificación a través de Internet, para lo cual deberá remitir a la Agencia la Hoja de solicitud correspondiente al envío realizado debidamente firmada. Por último, puede optar por el modo de presentación en soporte papel.*



## Auditoría de calidad de bases de datos de carácter personal

*Así mismo, permite notificar de forma simplificada, los ficheros de titularidad privada de comunidades de propietarios, clientes, libro recetario de las oficinas de farmacia, pacientes, gestión escolar, nóminas, recursos humanos y video vigilancia y los de titularidad pública de recursos humanos, gestión del padrón, gestión económica o control de acceso. A través de esta opción, el formulario electrónico muestra una notificación pre cumplimentada que podrá completar, o en su caso, adaptar a la situación concreta del fichero a notificar. En el caso de que ninguna de las notificaciones tipo previstas por la AEPD se adapte al fichero que pretende notificar, podrá seleccionar la opción de notificación normal.*

*Además, para los responsables que utilicen sus propios programas informáticos o los desarrolladores de aplicativos de protección de datos, se encuentran disponibles los formatos y especificaciones XML que deben cumplir sus aplicaciones para enviar notificaciones a la AEPD.”*

Pero, ¿quién debe notificar?

### Quiénes deben notificar

*“Están obligados a notificar la creación de ficheros para su inscripción en el RGPD, de acuerdo con lo dispuesto en la Ley Orgánica 15/99, aquellas personas físicas o jurídicas, de naturaleza pública o privada, u órgano administrativo, que procedan a la creación de ficheros que contengan datos de carácter personal.*

*La creación, modificación o supresión de los ficheros de las Administraciones Públicas sólo podrán hacerse por medio de disposición general publicada en el "Boletín Oficial del Estado" o diario oficial correspondiente.*

*Podrán crearse ficheros de titularidad privada que contengan datos de carácter personal cuando resulte necesario para el logro de la actividad u objeto legítimos de la persona, empresa o entidad titular y se respeten las garantías establecidas en la LOPD.*

*Toda persona o entidad que proceda a la creación de ficheros de datos de carácter personal lo notificará previamente a la Agencia Española de Protección de Datos.*

*Cualquier modificación posterior en el contenido de la inscripción de un fichero en el RGPD, deberá comunicarse a la Agencia Española de Protección de Datos, mediante una solicitud de modificación o de supresión de la inscripción, según corresponda. En ambos casos será necesario citar el Código de Inscripción asignado por el RGPD al fichero.*



## Auditoría de calidad de bases de datos de carácter personal

*En el supuesto de ficheros y tratamientos no automatizados, creados con posterioridad a la fecha de entrada en vigor de la LOPD (14 de enero de 2000), los mismos deberán ser notificados para su inscripción en el RGPD.*

*La no notificación de la existencia de un fichero supondría una infracción leve o grave, tal y como señala el art. 44 de la LOPD, quedando sujeto al régimen sancionador previsto en esta Ley.”*

### Pasos a seguir para cumplimentar el fichero NOTA

Para empezar, el responsable de notificar la creación, modificación o supresión de ficheros deberá descargarse el fichero de la página de la AEPD. Este fichero es totalmente gratuito.

Como se puede observar en la siguiente imagen, el fichero será distinto para cada una de las tres acciones a efectuar sobre los ficheros:

AGENCIA ESPAÑOLA DE PROTECCIÓN DE DATOS

Fichero de titularidad pública  
CONTENIDO DE LA NOTIFICACIÓN

NOTIFICATIONES ELEMATICAS A LA AEPD

**Tipo de solicitud de inscripción.**

Indique qué operación va a realizar sobre el fichero. En caso de modificaciones y supresiones se deberá indicar el Código de Inscripción que se asignó al fichero en el momento de su alta en el RGPD así como el CIF/NIF con el que fue inscrito. En caso de modificación se solicitan los apartados que se desea modificar y el Nombre o Razón Social del responsable.

☐ Alta ☐ Modificación ☐ Supresión

**Ilustración 57. NOTA 1**

Una vez seleccionada la acción, el responsable deberá ir cumplimentando el formulario entero. Esta cumplimentación la podrá realizar mediante papel (documento físico), o en Internet, firmando mediante un certificado digital.



## Auditoría de calidad de bases de datos de carácter personal

**Presentación de la documentación**

¿Cuál es el sistema que empleará para presentar la declaración?

☐ Formulario en papel

☐ Internet

☐ Internet firmado con certificado digital

**Ilustración 58. NOTA 2**

Dependiendo de la acción seleccionada por el usuario final, le saldrán unas opciones u otras:

- Tipo de solicitud de inscripción – Alta.

**Tipo de solicitud de inscripción.**

Indique qué operación va a realizar sobre el fichero. En caso de modificaciones y supresiones se deberá indicar el Código de Inscripción que se asignó al fichero en el momento de su alta en el RGPD así como el CIF/NIF con el que fue inscrito. En caso de modificación se solicitan los apartados que se desea modificar y el Nombre o Razón Social del responsable.

☒ Alta ☐ Modificación ☐ Supresión

**Modelo de declaración**

Si la notificación se refiere a un tratamiento de datos sobre miembros de comunidades de propietarios, clientes propios, libro de recetario, (clientes de farmacias), nóminas-recursos humanos (empleados) o pacientes, y la finalidad es la gestión propia de estos colectivos, puede marcar el cuadro TIPO y seleccionar el modelo que corresponda (se rellenan determinados apartados con valores apropiados) o bien seleccionar NORMAL para partir de un formulario totalmente vacío.

☒ Normal ☐ Tipo

**Ilustración 59. NOTA 3**

- Tipo de solicitud de inscripción – Modificación.



## Auditoría de calidad de bases de datos de carácter personal

**Tipo de solicitud de inscripción.**

Indique qué operación va a realizar sobre el fichero. En caso de modificaciones y supresiones se deberá indicar el Código de Inscripción que se asignó al fichero en el momento de su alta en el RGPD así como el CIF/NIF con el que fue inscrito. En caso de modificación se solicitan los apartados que se desea modificar y el Nombre o Razón Social del responsable.

☐ Alta ☒ Modificación ☐ Supresión

**Modificación**

Denominación social del responsable del fichero

CIF/NIF con el que figure inscrito el fichero

Códigos de Inscripción (Puede cumplimentar hasta 12 códigos de inscripción separándolos por punto y coma)

Al marcar la casilla correspondiente a "Supr" se suprimirá el apartado para el/los fichero/s seleccionados.

Supr. Mod.	Supr. Mod.
<input type="checkbox"/> Responsable del fichero o tratamiento	<input type="checkbox"/> Tipos de datos, estructura y organización del fichero
<input type="checkbox"/> Derechos de oposición, acceso, rectificación y cancelación	<input type="checkbox"/> Medidas de seguridad
<input type="checkbox"/> Encargado del tratamiento	<input type="checkbox"/> Cesión o comunicación de datos
<input type="checkbox"/> Identificación y finalidad del fichero	<input type="checkbox"/> Transferencias internacionales
<input type="checkbox"/> Origen y procedencia de los datos	

### Ilustración 60. NOTA 4

- Tipo de solicitud de inscripción – Supresión.

**Tipo de solicitud de inscripción.**

Indique qué operación va a realizar sobre el fichero. En caso de modificaciones y supresiones se deberá indicar el Código de Inscripción que se asignó al fichero en el momento de su alta en el RGPD así como el CIF/NIF con el que fue inscrito. En caso de modificación se solicitan los apartados que se desea modificar y el Nombre o Razón Social del responsable.

☐ Alta ☐ Modificación ☒ Supresión

**Supresión de la inscripción del fichero**

Denominación social del responsable del fichero

CIF/NIF con el que figure inscrito el fichero

Códigos de Inscripción

### Ilustración 61. NOTA 5

## Resumen estadísticas mensuales RGPD

*Datos recogidos de la AEPD del mes de Diciembre del año 2012.*

A continuación se mostrará una serie de resúmenes de las operaciones de inscripción, consulta y descarga de ficheros en el Registro General de Protección de Datos de la Agencia Española de Protección de Datos. Estos serán un resumen de operaciones de inscripción (altas, modificaciones y supresiones), un resumen por titularidad (pública y privada), un resumen de las consultas efectuadas a los registros, y las descargas efectuadas del registro general de protección de datos.

- Respecto a las operaciones de inscripción recogidas, podemos ver que el número de altas es mucho más elevado que el número de modificaciones y bajas, el cuál permanece bastante equiparado. Podemos observar, que en este mes no ha existido ninguna subida en cuanto al número de operaciones respecto del año en general, sigue una media bastante clara.

### 1. RESUMEN DE OPERACIONES DE INSCRIPCION

Tipo de operación	DICIEMBRE 2012	Total acumulado 2012
A INSTANCIA DEL INTERESADO		
Altas	29.356	450.046
Modificaciones	4.213	66.130
Supresiones	3.139	56.401
<b>TOTAL</b>	<b>36.708</b>	<b>572.577</b>
DE OFICIO		
Correcciones de oficio	1.274	57.674
<b>TOTAL OPERACIONES</b>	<b>37.982</b>	<b>630.251</b>

Media diaria de operaciones: 1.899

### Ilustración 62. NOTA 6

- Estas inscripciones se pueden dividir según su titularidad, privada o pública. Como podemos ver, la inscripción de ficheros por titularidad privada es mucho más elevado que el número de ficheros por titularidad pública (se muestra un resumen total):





**DISTRIBUCIÓN POR TITULARIDAD**

Titularidad	Nº de ficheros
Titularidad Pública	137.396
Titularidad Privada	2.865.720
<b>Total</b>	<b>3.003.116</b>

**Ilustración 63. NOTA 7**

- En lo referente a las consultas efectuadas a los registros generales de protección de datos, podemos comprobar, como el número de inscripciones por titularidad, que se efectúan algunas consultas más a ficheros de titularidad privada que a ficheros de titularidad pública:

**CONSULTAS AL REGISTRO GENERAL DE PROTECCIÓN DE DATOS**

	DICIEMBRE 2012	Total acumulado 2012
Titularidad Privada	220.174	3.380.914
Titularidad Pública	204.445	1.266.784

**Ilustración 64. NOTA 8**

- Por último, haciendo referencia a la descarga del formulario NOTA, podemos comprobar, como estamos viendo en todos los resúmenes, que existen un número de descargas de formularios para titularidad privada, que para titularidad pública:

**DESCARGAS DEL REGISTRO GENERAL DE PROTECCIÓN DE DATOS**

		DICIEMBRE 2012	Total acumulado 2012
Formulario NOTA	Titularidad Privada	23.122	334.688
	Titularidad Pública	2.044	27.551
Guía-modelo de Documento de Seguridad		9.557	99.264

**Ilustración 65. NOTA 9**

## **Resumen resoluciones LOPD**

A continuación se muestran resumidas 4 resoluciones extraídas de la AEPD, una referente al artículo 4.1 otra al 4.2, otra al 4.3 y otra al 4.4, existentes en el registro de resoluciones de la LOPD (para rescatar la resolución entera, consultar la página de la AEPD:

<https://www.agpd.es/portalwebAGPD/resoluciones/index-ides-idphp.php>).

### **RESOLUCIÓN: R/01512/2010.**

Un usuario denuncia a la Sociedad Estatal de Correos y Telégrafos por adjuntar unos datos personales, como son el nombre, los apellidos, el domicilio, la localidad de residencia y el teléfono del particular, a una evaluación de Seguridad y Salud de la empresa en la que trabaja, en la que supone que únicamente debería constar el puesto de trabajo del empleado.

La AEPD, haciendo referencia al artículo 4.1, entre otros artículos, dictamina que la sociedad cometió un error al adjuntar, los datos referentes al domicilio particular y teléfono de contacto, cuando no lo tenía que haber hecho, por lo que le notifica a las dos partes la sanción correspondiente debido a la infracción del artículo 4 de la LOPD.

### **RESOLUCIÓN: R/00875/2010**

Un usuario denuncia a la Diputación de Castellón el hecho de haberle enviado al domicilio de trabajo una multa de tráfico, sin intentarlo si quiera en el domicilio particular.

La AEPD, haciendo referencia al artículo 4.2, dictamina que la diputación no puede utilizar los datos de carácter personal para otra finalidad con la que fueron recogidos, por lo que notifica a las dos partes la sanción correspondiente debido a la infracción del artículo 4 de la LOPD.

### **RESOLUCIÓN: R/01760/2010**

Un usuario denuncia a la empresa France Telecom España por no haber actualizado los registros referentes al domicilio particular del titular del contrato de telefonía móvil, lo que ha conllevado a sumar una deuda de la que inicialmente no tenían información alguna.

La AEPD, haciendo referencia al artículo 4.3, dictamina que la empresa no ha puesto al día los datos de carácter personal de modo que no corresponden con



Auditoría de calidad de bases de datos de carácter personal  
veracidad la situación actual del afectado, por lo que notifica a las dos partes la sanción correspondiente debido a la infracción del artículo 4 de la LOPD.

**RESOLUCIÓN: R/00091/2011**

Un usuario denuncia a la Universidad Europea de Madrid por haber facilitado datos personales suyos para notificar a la Jefatura Provincial de Tráfico de Madrid una sanción con un coche de empresa que seguía estando a su nombre, cuando él ya no era parte de dicha universidad.

La AEPD, haciendo referencia a los artículos 4.3 y 4.4, dictamina que los datos de carácter personal no se han actualizado, y que los datos resultan ser inexactos, todos o en parte o, incompletos, y no se han cancelado ni sustituido por otros datos, por lo que notifica a las dos partes la sanción correspondiente debido a la infracción del artículo 4 de la LOPD.



## ***Ponderaciones de los campos de las diferentes bases de datos***

En esta sección se explica la ponderación puesta a cada uno de los campos de las cuatro diferentes bases de datos:

- **Universidad:**
  - Campos de identificación: se tomará con más importancia los campos referentes a los apellidos, pues serán los identificativos dentro de esta bbdd.
  - Campos de dirección: en referencia a estos campos, el campo más significativo será el del email, pues es una herramienta primaria de comunicación entre profesores y alumnos por ejemplo.
- **Banco:**
  - Campos de identificación: en esta bbdd el campo con más peso será el que contiene el DNI de la persona, y se dejará con menos peso, o importancia a los referentes a la fecha de nacimiento del usuario.
  - Campos de dirección: se tomará con más relevancia los campos referentes a la calle y población del usuario y no tanto al teléfono móvil y menos al email de la persona, pues es un elemento no primario en la comunicación con el banco.
- **Web de empleo:**
  - Campos de identificación: igual que en la bbdd del banco el campo o atributo con mayor relevancia será el DNI, acompañado del nombre y apellidos de la persona, pues son los más identificativos en relación a su búsqueda de empleo.
  - Campos de dirección: en dicha ponderación toma mucho peso los atributos del teléfono móvil y el email de la persona, pues son los elementos más utilizados en este tipo de bbdd.
- **Videoclub:**
  - Campos de identificación: por último, en esta bbdd el atributo con mayor ponderación será el nombre y los apellidos pues serán los que identifique a cada usuario del videoclub, dejando sin mayor relevancia los campos de sexo o edad de la persona.
  - Campos de dirección: en este caso el atributo con mayor peso será el del teléfono móvil, pues será utilizado por el personal del videoclub para cualquier incidencia de cualquier tipo con el usuario.

*(Cabe recordar que cada ponderación, como se ha mencionado con anterioridad, se podrá modificar previa petición del propietario de dicho servicio web)*



Universidad  
Carlos III de Madrid